



# COGNITIVE DIFFERENCES & REPETITION INFLUENCES ONE'S SUSCEPTIBILITY TO THE ILLUSORY TRUTH EFFECT

By: Kyla Allan, Anna Cole, Rebecca Ferguson, Olivia Holm



## INTRODUCTION

**Illusory Truth Effect:** the tendency to perceive repeated statements as more believable even when they are false [1].

**Need for cognition:** a cognitive trait which describes the tendency to engage in effortful thinking. Individuals with high NFC tend to engage in more effortful thinking than those with low NFC.

**Research Question:** What is the influence of cognitive abilities on people's susceptibility to believing false truths?

**Hypothesis:** NFC will moderate the ITE

## METHODOLOGY

### Participants

- N= 273, English speaking Canadians
- Ages = 18-26



### Materials

- Statements for truth rating [2]
- e.g. "Lima is the capital of Chile"
- NFC assessment [3]
- e.g. 'I would prefer complex to simple problems'

### Procedure

- Interest rating on trivia questions
- Truth rating on trivia questions
- Complete nfc questionnaire

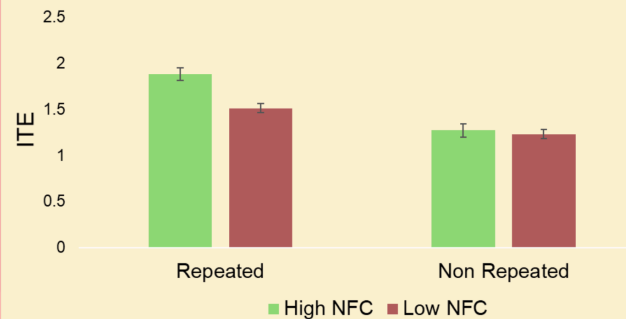
## FINDINGS

There was a main effect of NFC, indicating a significant difference between participants with high and low NFC

There was no main effect of repetition on ITE after controlling for both age and gender of participants

A significant interaction was found between repetition and NFC groups. This means that the effect of repetition varied based on NFC group.

Mean scores of the ITE for repetition and NFC



## IMPLICATIONS

These findings suggest high-NFC individuals process information more deeply, increasing fluency processes. These results could be used to inform media-literacy education, and future research on the complex relationship between repetition and belief, and the effect of individual differences.

## CONCLUSION

**NFC and ITE:** Important role in cognitive pathways relating to how people process information and may offer some explanation as to why some people are more susceptible to repeated falsehoods than others.

**Repetition:** Our results show a complex relationship, indicating that its effects on ITE are not as straightforward as they may seem. Instead, some demographic characteristics, including age and gender, may play a unique role



## REFERENCES

- [1] Fazio, L. K. (2020). Repetition Increases Perceived Truth Even for Known Falsehoods. *Collabra: Psychology*, 6 (1): 38. doi: Fazio, L. K. (2020). Repetition Increases Perceived Truth Even for Known Falsehoods. *Collabra, Psychology* (2020) 6 (1): 38.
- [2] Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.
- [3] Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms; updated and expanded from the Nelson and Narens. *Behavioural Research*, 45:11115-1143.



Department of Cognitive Science

# ALLOCATION OF COGNITIVE EFFORT IN DEPRESSION DURING NATURALISTIC READING

## Authors

Akshaya Kirithy Baskar, Nadine Charanek, Olessia Jouravlev



## Affiliations

Department of Cognitive Science, Carleton University



## Introduction

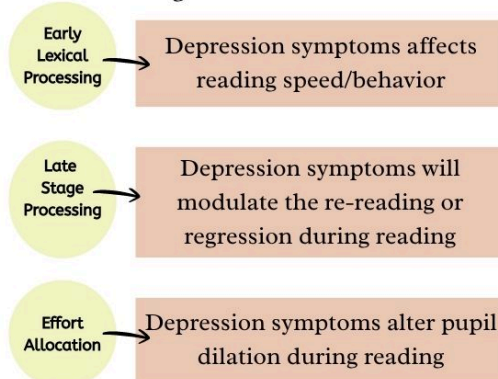
Major Depressive Disorder has been linked to alterations in attention, cognitive control, and motivation (Pizzagali, 2014), yet its impact on naturalistic language processing remains unexplored. To date, research on depression and language has focused primarily on offline measures or simplified experimental paradigms (Rayner, 2009).

## Novelty

This study leverages eye-tracking in a naturalistic reading context to capture continuous processing dynamics (Futrell et al., 2018), offering a novel approach to disentangling whether observed differences reflect core linguistic deficits or altered effort allocation (Fedorenko et al., 2013).

## Hypothesis

**Core Question**  
Does depression alters the language processing or allocation of cognitive effort?



## Methodology

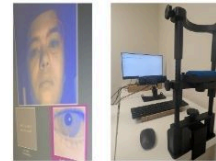
66 participants, Mean age = 21.65

### Online Session

- Beck Depression Index (*Depression*)
- Peabody Picture Vocabulary Test, Test for Receptive of Grammar (*Language*)
- Kaufman Brief Intelligence Test (*IQ*)

### In person Session

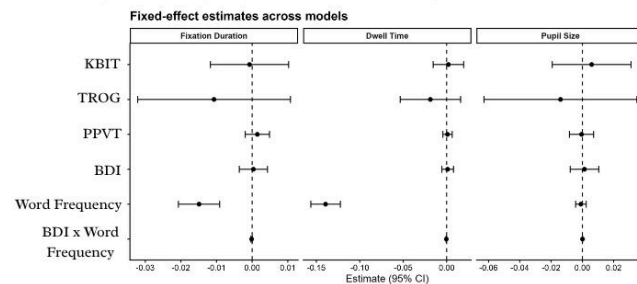
Eye tracking testing short stories from fiction (Futrell et al)



For data analysis we used SR Research Data Viewer and R statistical software for linear mixed modelling.

## Results

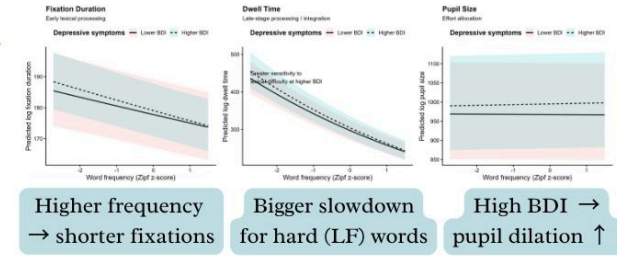
Participant level analysis didn't show any changes in fixation time, re-reading pattern, or pupil dilation. In word level analysis models of rereading revealed no significant effects of BDI ( $\beta = 0.002$ ,  $SE = 0.006$ ,  $p = .73$ ). Variance showed that rereading behavior was primarily driven by word-level variability rather than participant.



Word frequency having strong negative effect.

BDI score did not impact fixation duration

BDI score x word frequency - significant in dwell time model.



## Discussion

### Altered Engagement Theory

- Early Lexical Processing: Depression symptoms affects reading speed/behavior
  - Late Stage Processing: Depression symptoms will modulate the re-reading or regression during reading
  - Effort Allocation: Depression symptoms alter pupil dilation during reading
- Early word recognition is intact. Depression increases sensitivity to linguistic difficulty during integration / reanalysis. Depression alters effort engagement on word frequency

High BDI may be exerting more effort overall, but not selectively for harder words. We controlled for language and cognitive ability to isolate depression-specific effects

## Future directions/Conclusion

- Word valence based modelling approach.
- Combining neural measures with eye tracking.
- Depression selectively alters when and how cognitive effort is deployed during reading.
- Overall, depression did not affect early lexical processing but modulated later-stage integration and effort allocation, particularly under increased linguistic difficulty.

## Key Sources & Acknowledgements

- The Natural Stories Corpus - Futrell et al. 2018
- Broad domain generality in focal regions of frontal and parietal cortex - Fedorenko et al., 2013
- Depression, stress, and anhedonia: toward a synthesis and integrated model - Pizzagali, 2014
- Eye movements and attention in reading, scene perception, and visual search - Rayner, 2009



Thanks to

Marina Panfilova, Nayna Kirubakar, Grace Yee, Fatemeh Maleki, Aahana Uppal, Amanda Schimdt, Naomi Brake

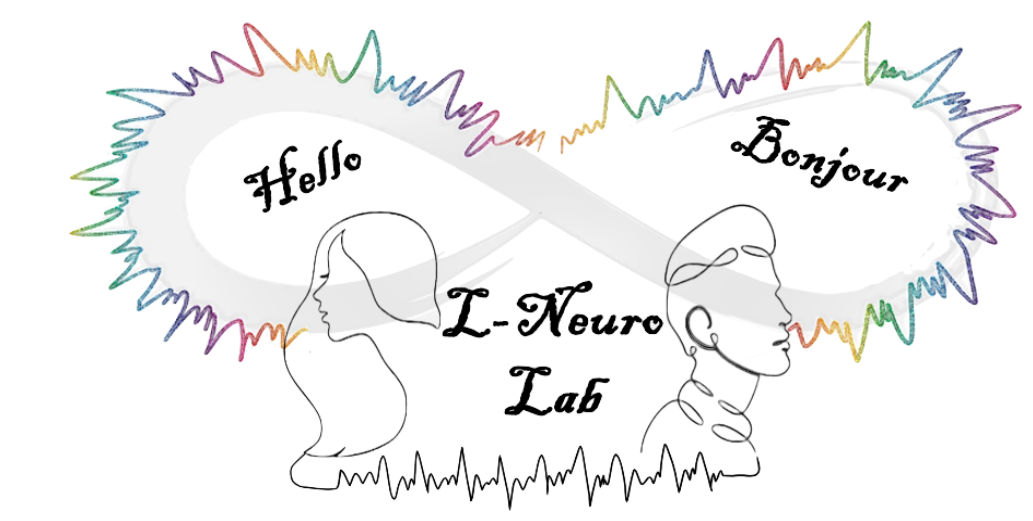


# Let's play telephone!

## Quantifying the effect of transparency bias on language evolution



Aya Amer, Olessia Jouravlev  
L-Neuro Lab, Department of Cognitive Science, Carleton University



### Background

#### Learning biases and language development

- Learning bias:
  - People make less mistakes with specific patterns
  - People are more likely to infer specific patterns from limited data
- Applied to language:
  - Learners introduce mistakes which regularize the language
  - When forced to extrapolate, learners create new utterances that adhere to their biases
  - Random communication system becomes "structured" – symbols systematically map to meanings

#### Proposed learning biases affecting language development

- Simplicity**
  - Bias toward patterns that minimize the number of words
  - Claim: learners' mistakes and inferences make language simpler
- Transparency**
  - Bias toward patterns that map every word to exactly one meaning
  - Claim: learners' mistakes and inferences make language more transparent

#### Iterated learning procedure

- Like a game of "telephone" (Figure 2)
- Participant responses on a vocabulary test are used to train next participant in the chain
- Used to investigate the influence of learning biases on language development

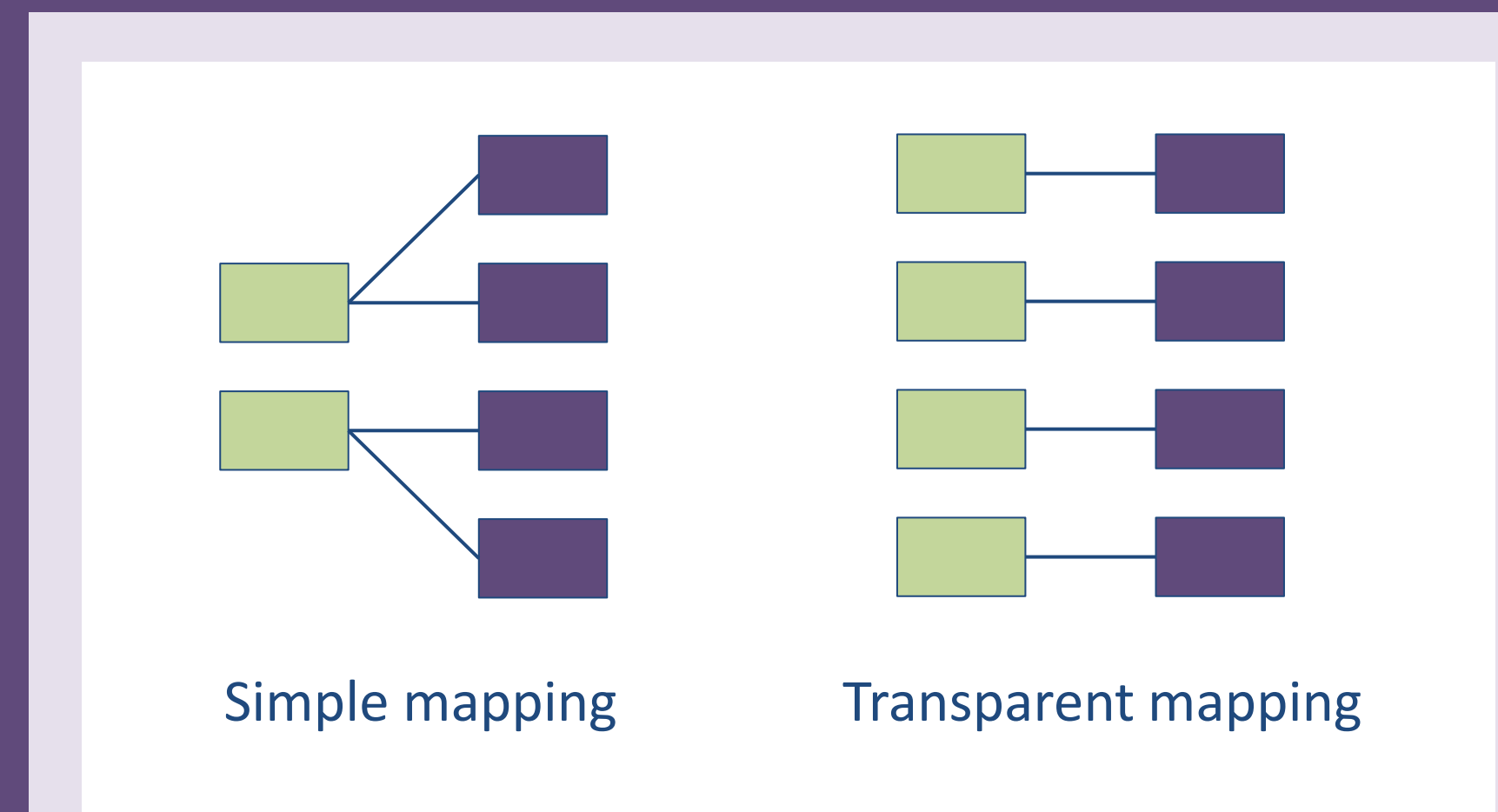


Figure 1: Mappings between words (green) and meanings (purple) reflecting the different proposed learning biases

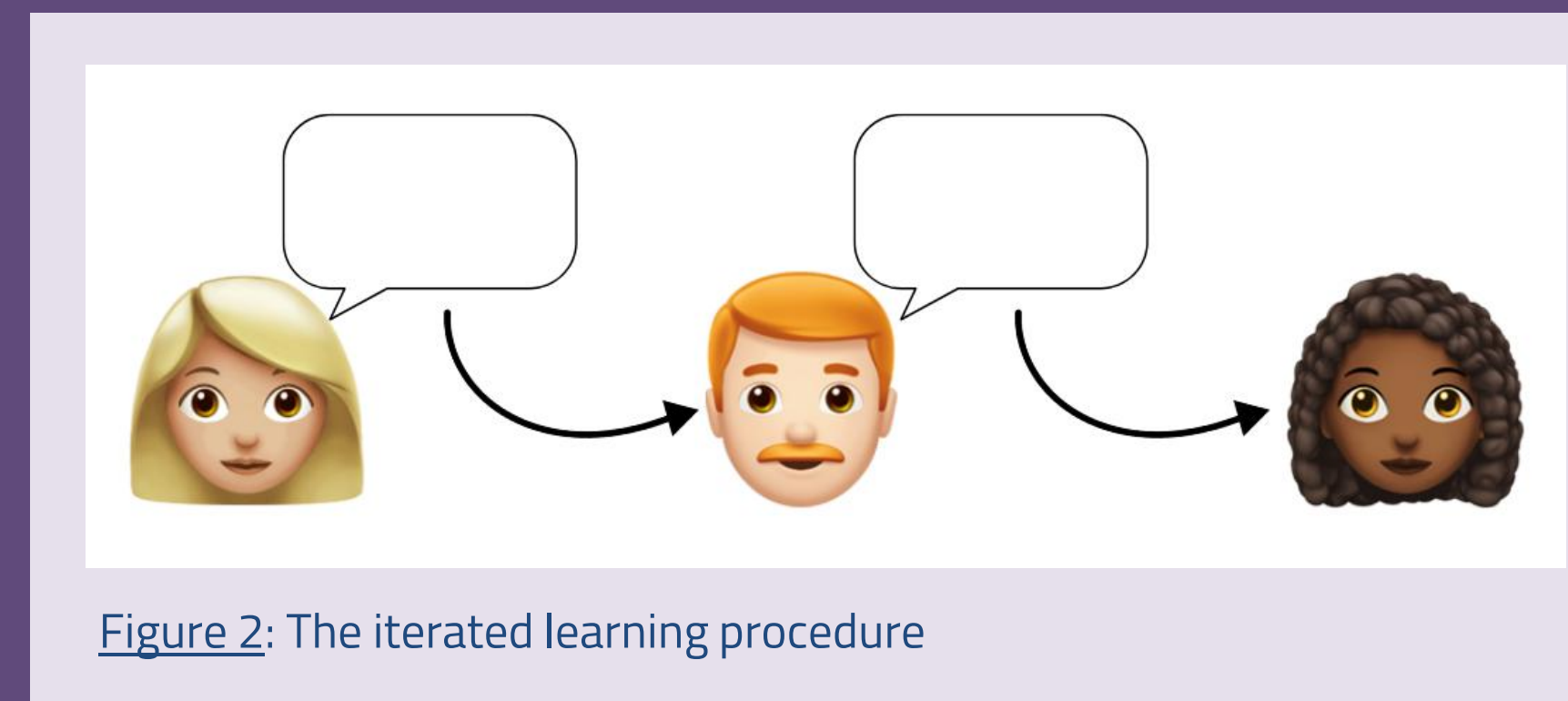


Figure 2: The iterated learning procedure

### Issues with Kirby et al.

- Participants influenced by their knowledge of English
- English has a low morpheme-to-word ratio
  - Turkish: Şimdi otobüs-te-yiz now bus-LOC-COP.1.PL
- English puts spaces between words
  - Japanese: この本は難しいです
  - this book TOP difficult COP
  - "This book is difficult"
- English speakers in Kirby et al.'s experiment see a one-word label and might assume that one word = one meaning
  - Would never produce compositional structure, because of language-specific biases

### Materials & Methods

#### Participants

- 200 participants (Figure 9)
- 10 transmission chains x 10 participants per chain x 2 experimental conditions



Figure 9: Example of 10 transmission chains with 10 participants each

#### Artificial language

- 27 items (Figure 10)
  - 3 shapes x 3 colours x 3 numbers
- Labels in Experiment 1: same length, no spaces
  - "wuneho"
  - "betoga"
  - "nukagi"
- Labels in Experiment 2: 1, 2, or 3 syllables, spaces
  - "wu ne ho"
  - "be tu"
  - "nu"

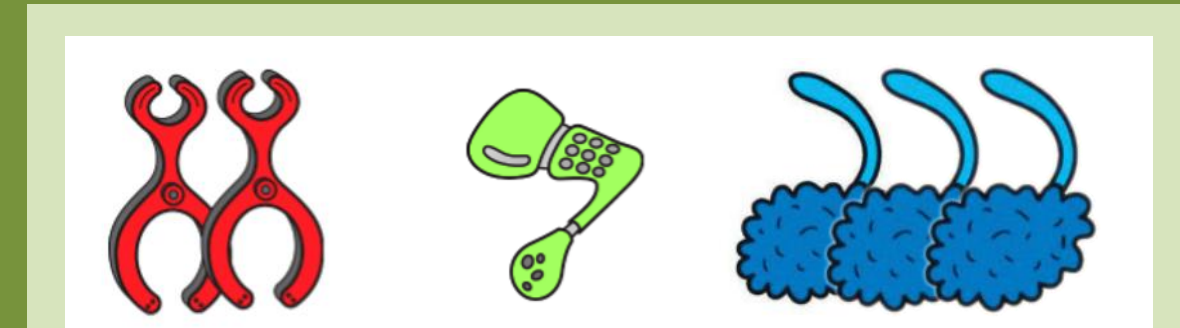


Figure 10: Example items (Beckner et al., 2017)

### Research Questions

- Aim: modify Kirby et al.'s experiment to control for English-specific learner biases
- Is the development of language through repeated learning influenced by simplicity bias or transparency bias?
  - Can we replicate Kirby et al.'s simplicity bias finding?
  - If we control for the confounds in Kirby et al., will languages develop transparent, compositional structure?

### Proposed Experiment

- Experiment 1: Replicate Kirby et al.
  - Label each stimulus with a single word
- Experiment 2: Label the stimuli with a variable number of syllables, separated by spaces
  - Allows participants to overcome the English-specific assumption that one word = one meaning
  - Variable number of syllables is to avoid biasing subjects toward compositionality

### Kirby et al.'s Experiment

#### Procedure

- Used iterated learning
- Taught artificial language to a participant, tested participant on the language, and used their answers to train the next participant
- Repeated for 10 participants per chain (4 chains total)
- Compared initial and final artificial languages

#### Prediction

- Artificial language would go from random structure (Figure 3&4) to compositional structure (Figure 5&6)
  - Because of memory demands, need to extrapolate

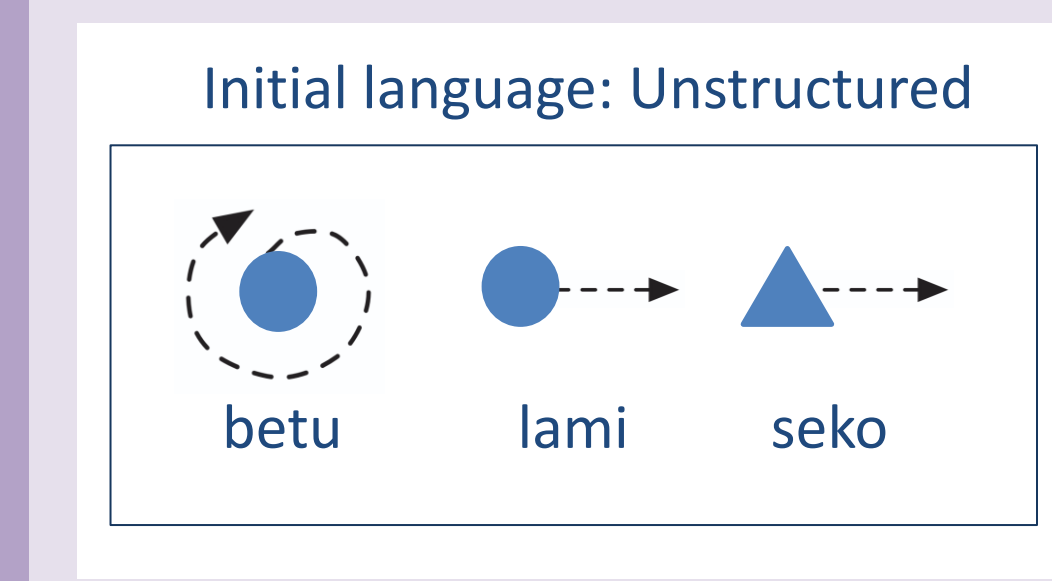


Figure 3: Initial structure of the artificial language

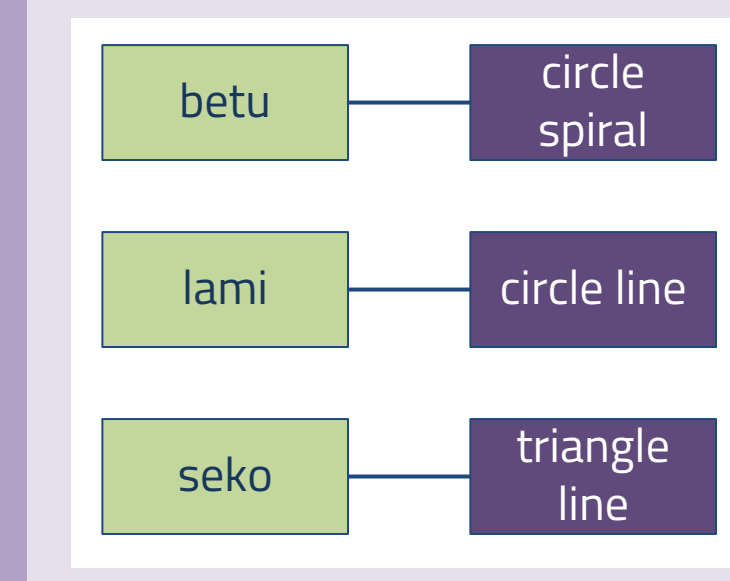


Figure 4: Initial word-meaning mapping

#### Results

- Artificial language became "degenerate": vocabulary cannot distinguish the full range of meanings (Figure 7&8)
- Concluded that learners show overwhelming simplicity bias

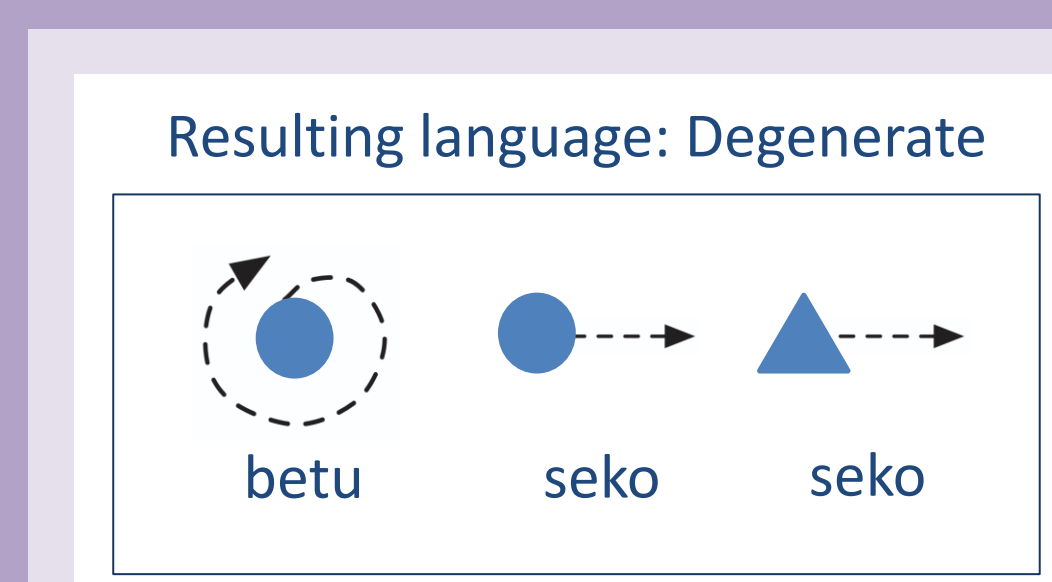


Figure 7: Structure of the artificial language at the end of the transmission chain

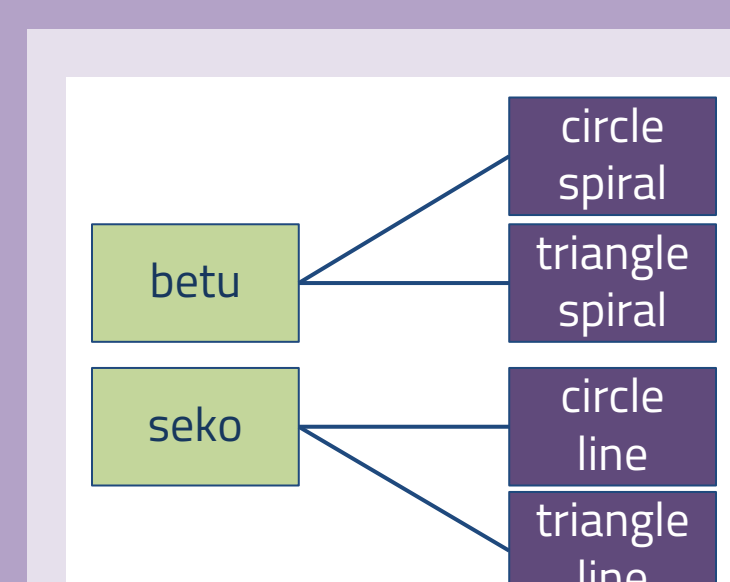


Figure 8: Final word-meaning mapping

#### Predicted language: Compositional

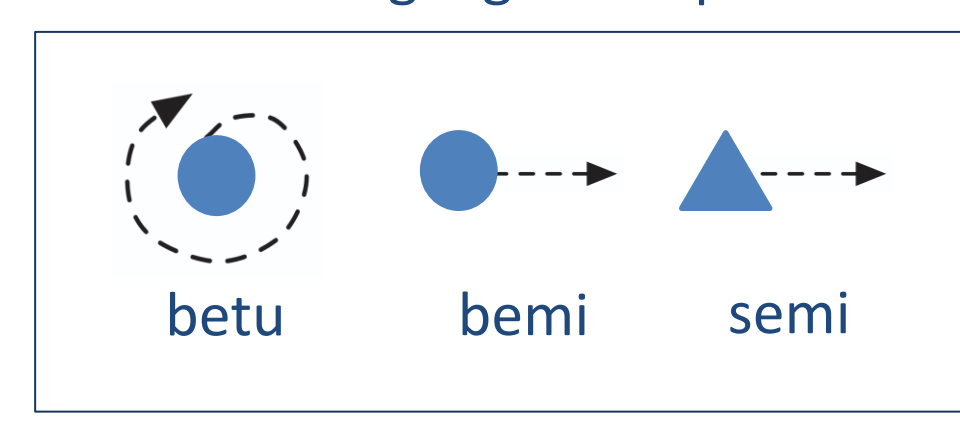


Figure 5: Prediction for the structure of the artificial language at the end of the chain

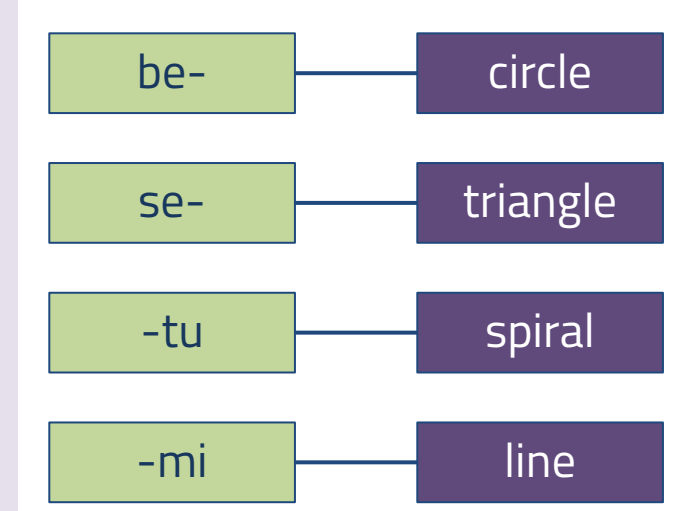


Figure 6: Predicted word-meaning mapping

### References

- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, 2(2), 160–176.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.

# Presupposition and Reasoning in Conditionals: A Theory-Based Study of Humans and LLMs

Tara Azin<sup>1</sup>, Yongan Yu<sup>2</sup>, Raj Singh<sup>1</sup>, Olessia Jouravlev<sup>1</sup>

<sup>1</sup>Carleton University, Department of Cognitive Science

<sup>2</sup>McGill University, School of Computer Science



## INTRODUCTION

Presupposition projection in conditional sentences sits at the intersection of semantics, pragmatics, and probabilistic reasoning. However, it remains largely unevaluated in large language models. This study presents a controlled behavioral comparison of human and LLM presupposition judgments on conditional sentences of the form *If A, Bp*, where the relationship between the antecedent and the projected presupposition is varied. Using a normed dataset of 90 items and likelihood ratings from 120 human participants and four LLMs, we investigate how antecedent-presupposition relevance and contextual information affect presupposition interpretation. Our findings show that while some models approximate human judgment patterns, this alignment appears to stem from surface-level pattern matching rather than real pragmatic competence.

## SETUP

We constructed 90 conditional sentences of the form *If A, Bp* across three antecedent-presupposition (A-p) relevance levels (relevant, somewhat relevant, and irrelevant) validated through a norming study with 30 native English speakers. Each sentence was paired with a target presupposition for likelihood rating.

**Participants & Models.** 120 native English speakers recruited via Prolific rated each item on a 0–7 Likert scale under two conditions: with and without minimal identifying context. The same stimuli and instructions were given to four LLMs (GPT-5, Gemini-2.5-flash, Llama-3.1-8B, and Qwen2.5-7B) which additionally produced chain-of-thought reasoning traces.

## SETUP

### Instructions

In this task, imagine you are participating in a get-to-know-people activity. Suppose that there are a bunch of cards with names on them, along with a small identifying detail about the person (e.g., where the person is from). This information is provided only to identify which person is being talked about.

Suppose that someone is tasked with selecting a card at random, and then -- depending on who the card identifies -- they have to say something they know about that person. Assume that the speaker is **honest, reliable, and helpful**, and that you share relevant background assumptions with them. You will then be asked to judge how likely a second statement is to be true on the assumption that what the speaker said is true. Please respond using a scale from 0 (very unlikely) to 7 (very likely).

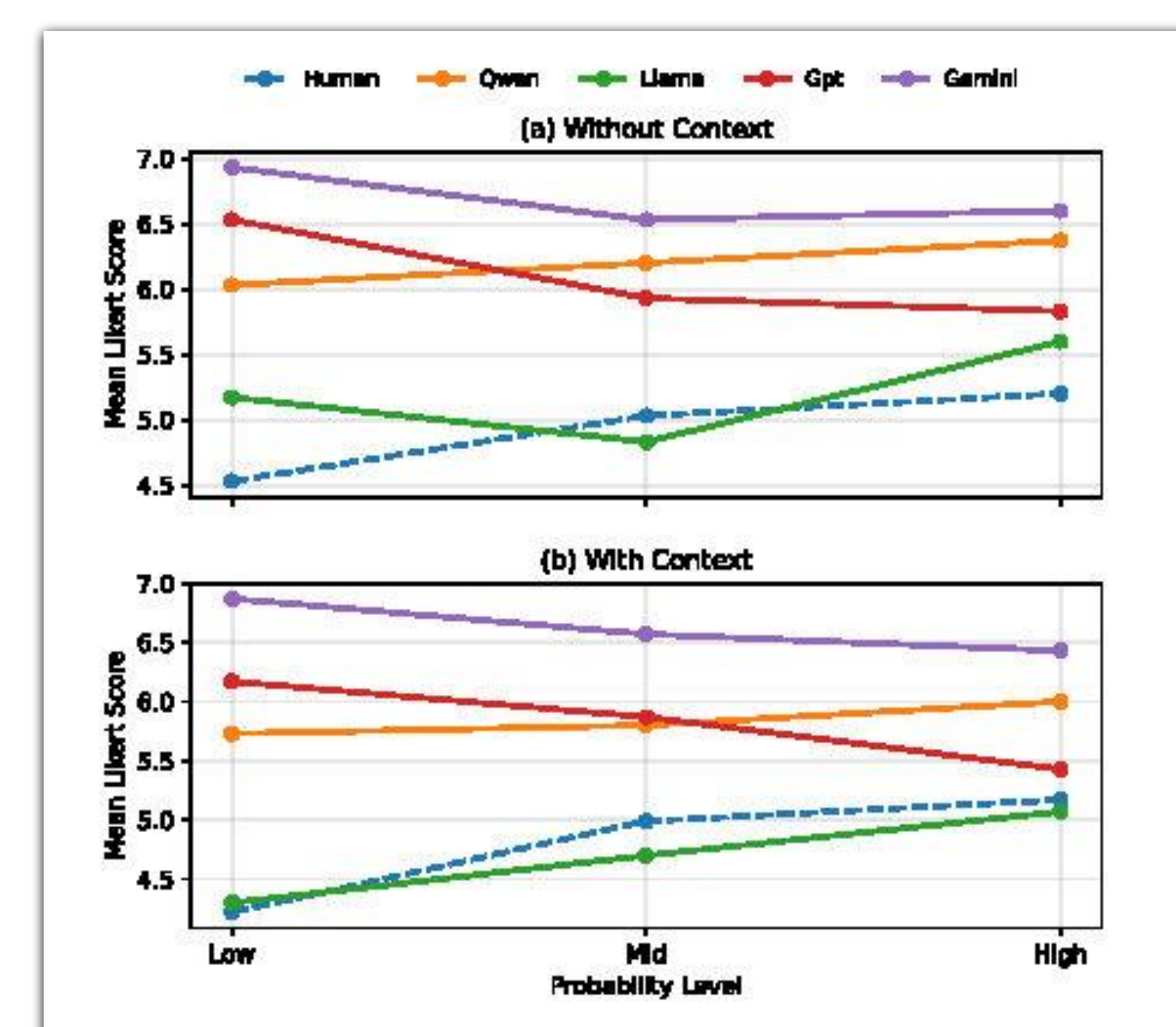
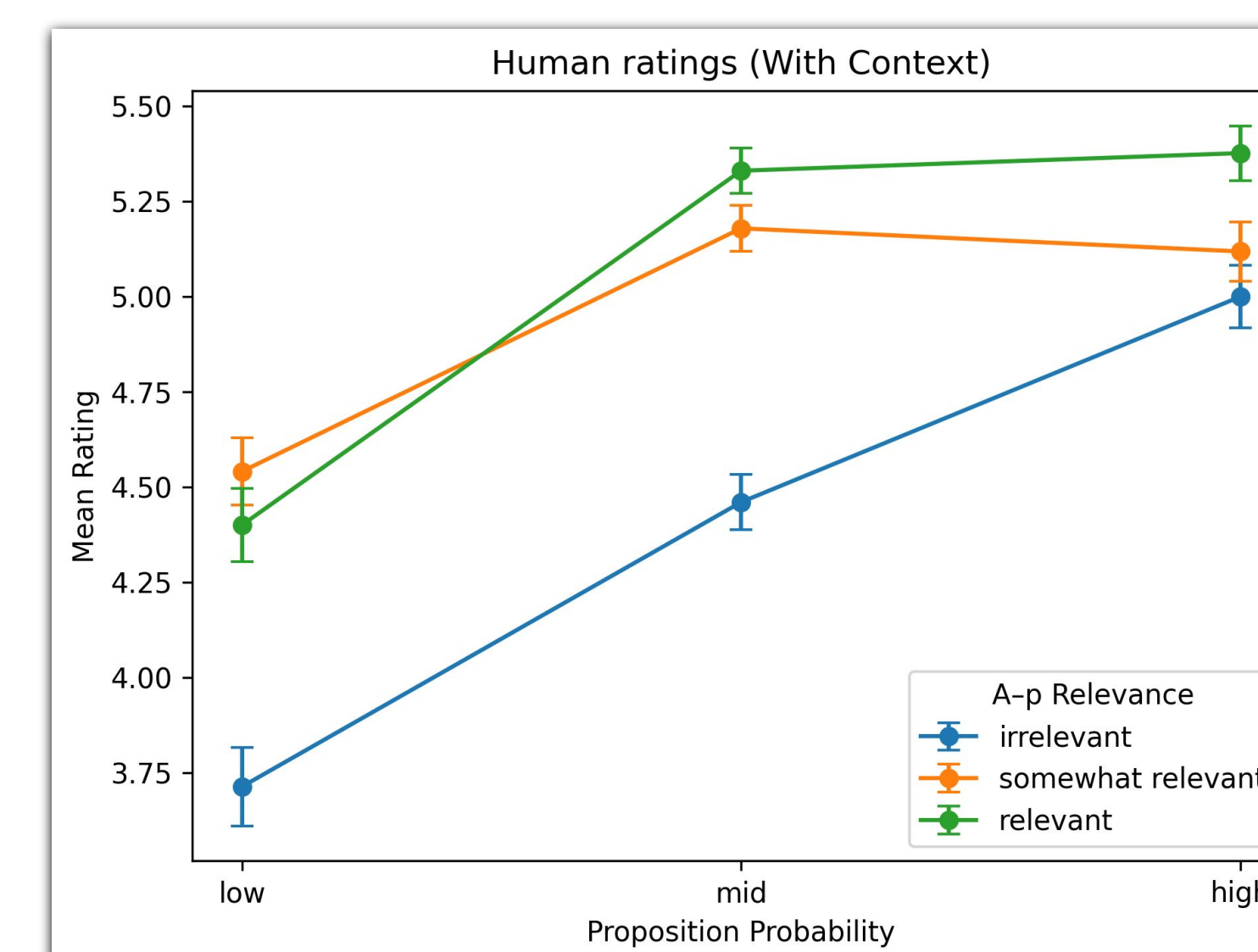
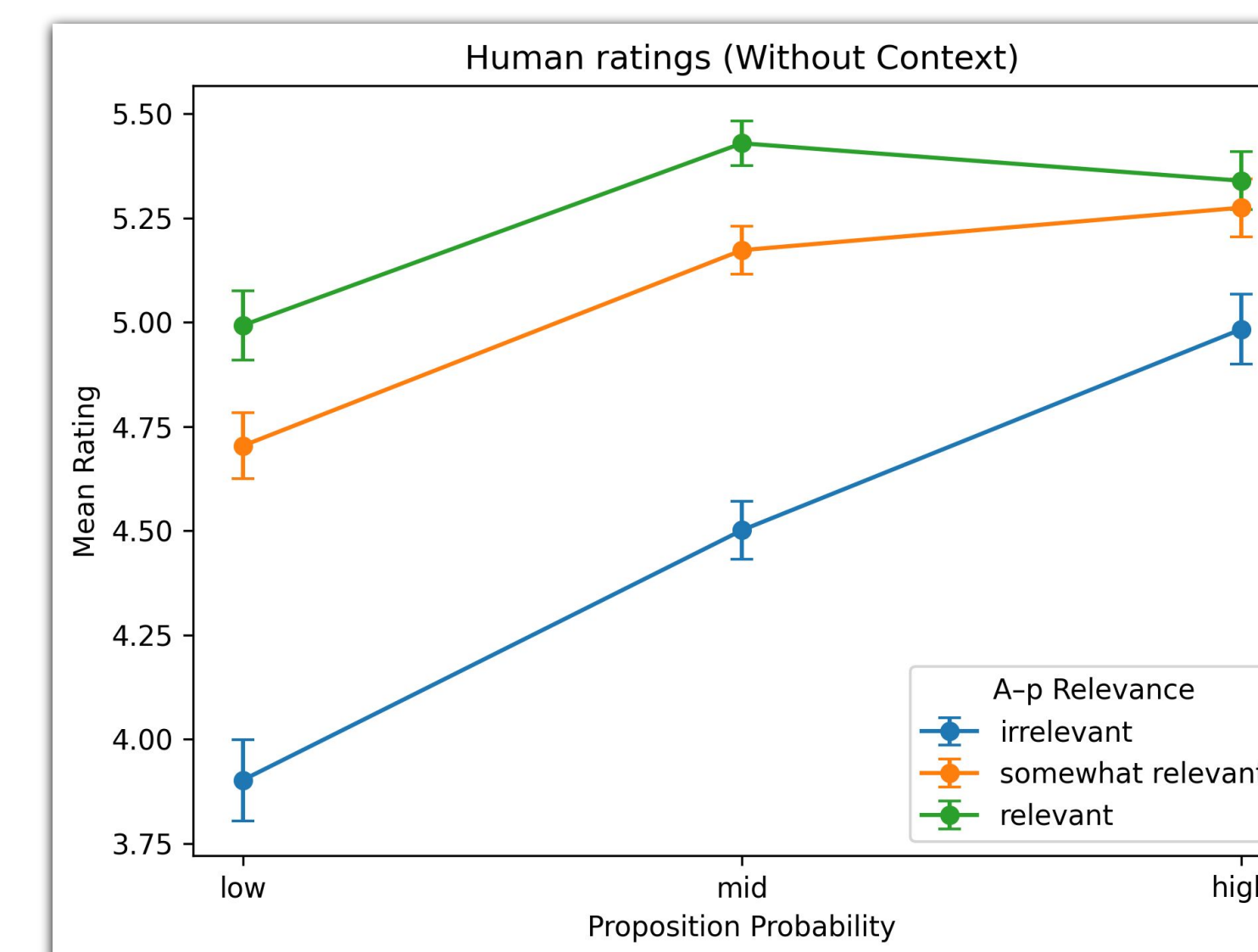
First, let's practice with two items. Please click "Next" to continue.



**Stimuli Design.** Base propositions spanned low- (e.g., *having a wetsuit*), mid- (e.g., *having a brother*), and high-probability (e.g., *having a smartphone*) ownerships, each embedded in a conditional with a possessive trigger. Antecedents were selected to induce graded A-p relevance, for instance, *being a scuba diver* is highly relevant to *having a wetsuit*, while *liking coffee* is irrelevant. This crossed design allowed us to examine how proposition probability and antecedent relevance jointly shape presupposition judgments.

**Reasoning Evaluation.** To assess whether model reasoning shows real pragmatic competence, we developed a theory-informed checklist grounded in formal semantics and pragmatics (covering accuracy, presupposition handling, pragmatic criteria, and coherence). A judge model (Claude Haiku 4) evaluated each reasoning trace against 52–59 binary checklist questions, with human annotators validating 5% of outputs (89% inter-annotator agreement).

## RESULTS



## RESULTS

Human judgments showed a clear interaction between proposition probability and A-p relevance, with irrelevant antecedents imposing strong penalties especially for low-probability items. Among models, Qwen2.5-7B and Llama-3.1-8B aligned most closely with human ratings, yet scored lowest on the reasoning checklist while proprietary models showed the opposite pattern. This dissociation suggests that behavioral alignment with humans shows surface-level pattern matching rather than true pragmatic competence.

## DISCUSSION

Our results highlight an important gap in LLM pragmatic reasoning. Matching human behavior and understanding language the way humans do are not the same thing. Models that best mimic human presupposition judgments do so through distributional pattern matching rather than principled pragmatic inference. This is a distinction that is invisible to accuracy-based metrics alone. Minimal context improved human sensitivity to probabilistic structure but often introduced inconsistencies in model outputs. These findings emphasize the need for theory-based evaluation frameworks that combine behavioral measures with reasoning assessment to meaningfully diagnose pragmatic competence in LLMs.

## REFERENCES

- I. Heim. "On the Projection Problem for Presuppositions". In: *Formal Semantics: The Essential Readings*. Ed. by P. Portner and B. H. Partee. Blackwell, 1983, pp. 249–260.
- T. Azin, D. Dumitrescu, D. Inkpen, and R. Singh. "Let's CONFER: A Dataset for Evaluating Natural Language Inference Models on CONditional INFERence and Presupposition". In: *Proceedings of the Canadian Conference on Artificial Intelligence*. 2025. doi: 10.21428/594757db.13f92bbd.
- S. Sravanthi, M. Doshi, P. Tankala, R. Murthy, R. Dabre, and P. Bhattacharyya. "PUB: A Pragmatics Understanding Benchmark for Assessing LLMs' Pragmatics Capabilities". In: *Findings of the Association for Computational Linguistics: ACL 2024*. Association for Computational Linguistics, 2024, pp. 12075–12097.

**Code tracing** is the process of simulating how a computer executes a program. This involves keeping track of variable values and state of program execution.

## Problem 1

```
total = 100
count = 1
scores = [10, 1, 13, 7, 3, 5, 4]
for val in scores:
    if val < 5:
        total = total - val
        count = count + 1
    print( count, total )
```

## Problem 4

```
cards = [5, 1, 10, 0, 9, 2]
result1 = 0
result2 = 1
final = 40
for element in cards:
    if element > 4:
        result1 = result1 + 1
    if element < 5:
        result2 = result2 + 2
    final = final - element
print(result1, result2, final)
```

## Code Tracing Strategies

### Assignment + Computation Sequence

Val = 10 > 5, count = 1 + 1 = 2, total = 100  
 Val = 1 < 5, total = 100 - 1 = 99  
 Val = 13 < 5, total = 99

### Condensed Calculation

*total = 100 - 1 - 3 - 4*

### Table

	1st	2nd	3rd	4th
val	10	1	13	7
count	1	2	3	4
total	100	99	96	92

### Cross Out

*total = 100 99 96 92*  
~~count = 1 2 3 4 5 6 7 8~~  
~~scores = [10, 1, 13, 7, 3, 5, 4]~~  
 val

### Other

A strategy is used which cannot fit into any of these categories and cannot be considered a combination of these strategies.

### None

No code trace is provided. Only an answer is written.

### Assignment Sequence

Line 1:  
total=100  
count=1

Line 2:  
total=100  
count=2

### Variable Value Sequence

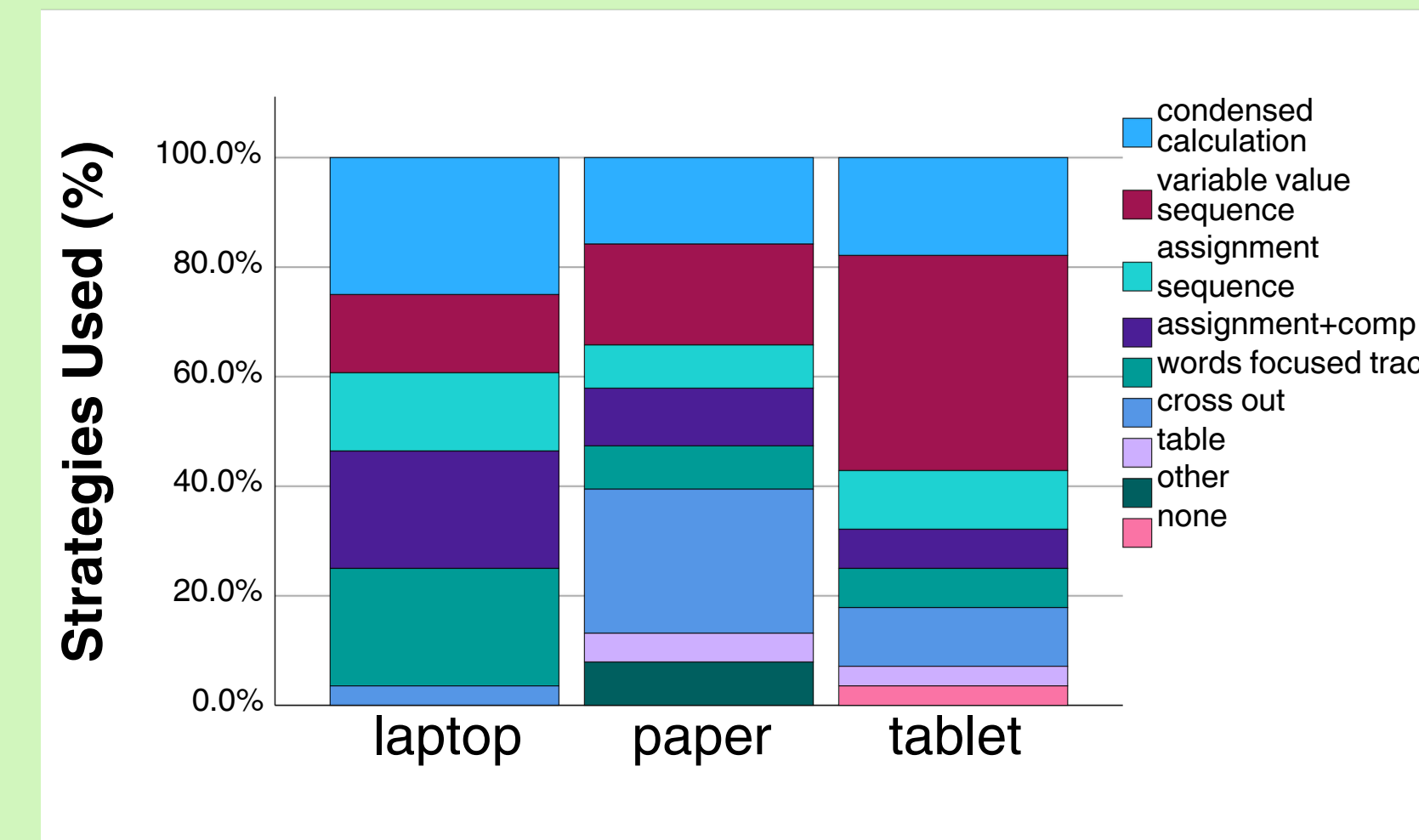
→ 2 100  
 → 3 99  
 → 4 100  
 → 5 100  
 → 6 97  
 → 7 100  
 → 8 96

### Words-Focused Trace

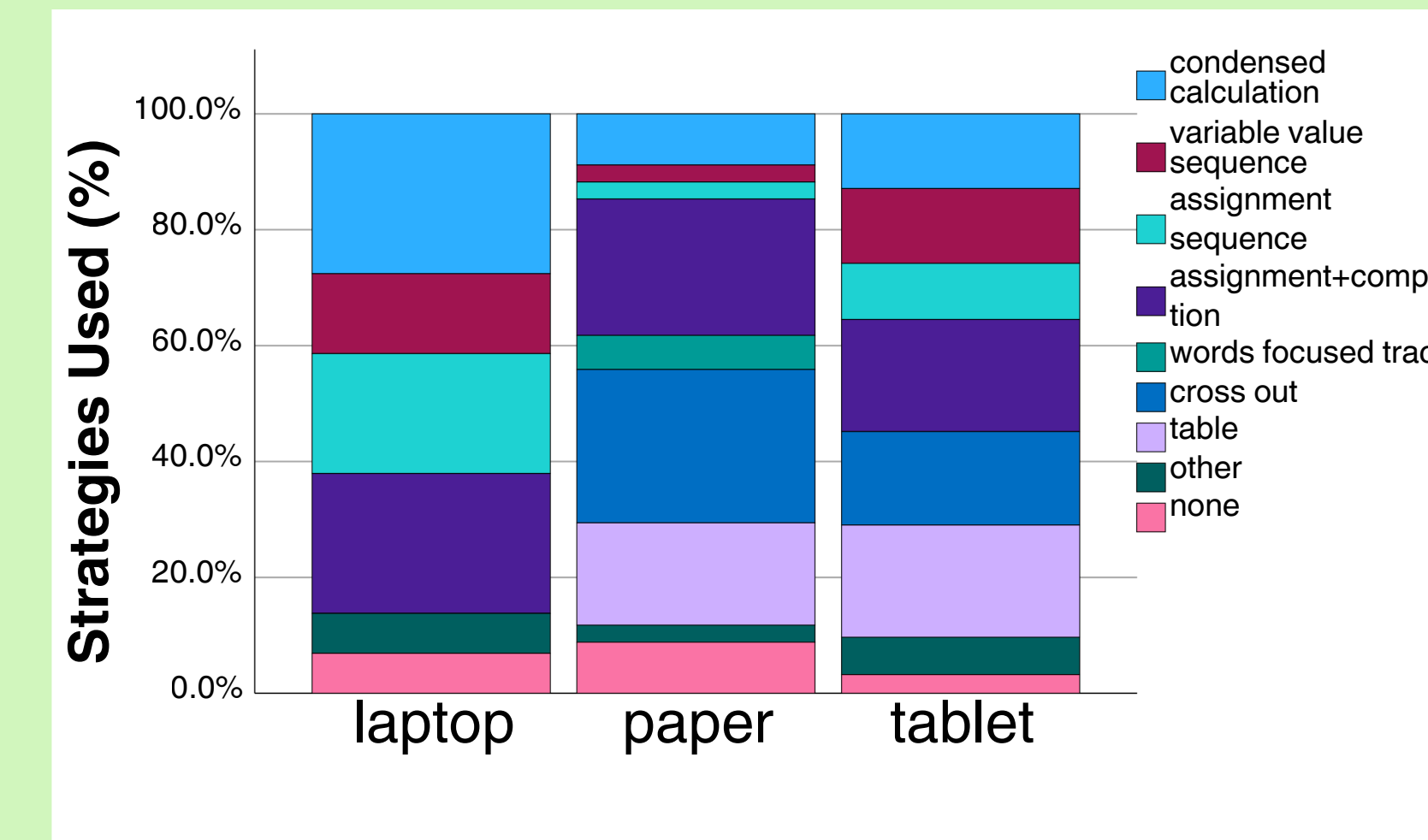
- total is assigned 100 (line 1)  
 - count is assigned 1 (line 2)  
 - scores is assigned [10, 1, 13, 7, 3, 5, 4] (line 3)  
 - for loop begins for scores (line 4)  
 - if value is smaller than 5 (line 5)  
 - ~~total~~ 100 is subtracted by the value (line 6)

## Proportion of Strategies Used by Modality

### Problem 1



### Problem 4



## Variables of Interest

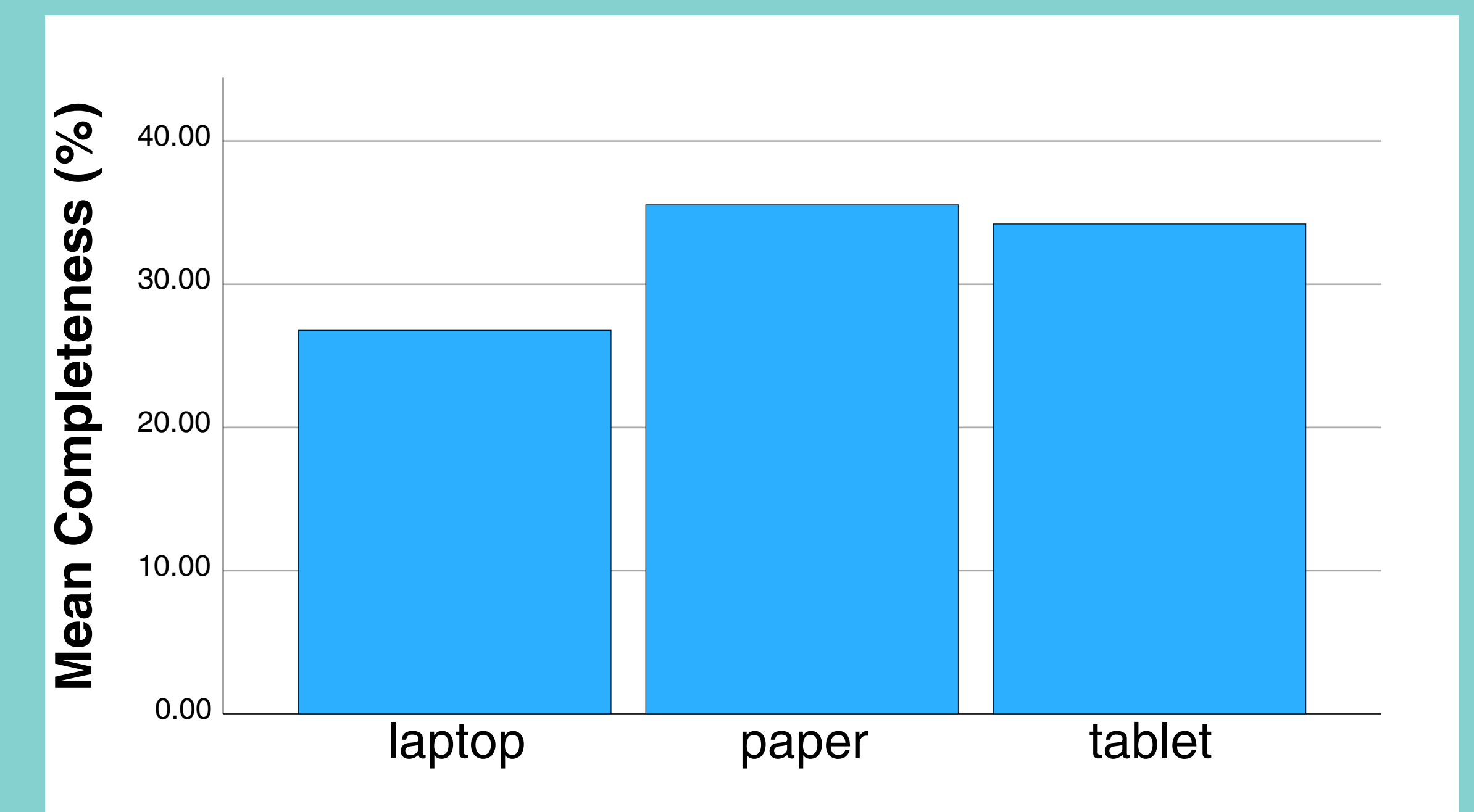
- Correctness:** number of printed variables containing a correct value

e.g. In problem 1, if the final output values of count and total are both correct, a score of 2/2 was assigned.

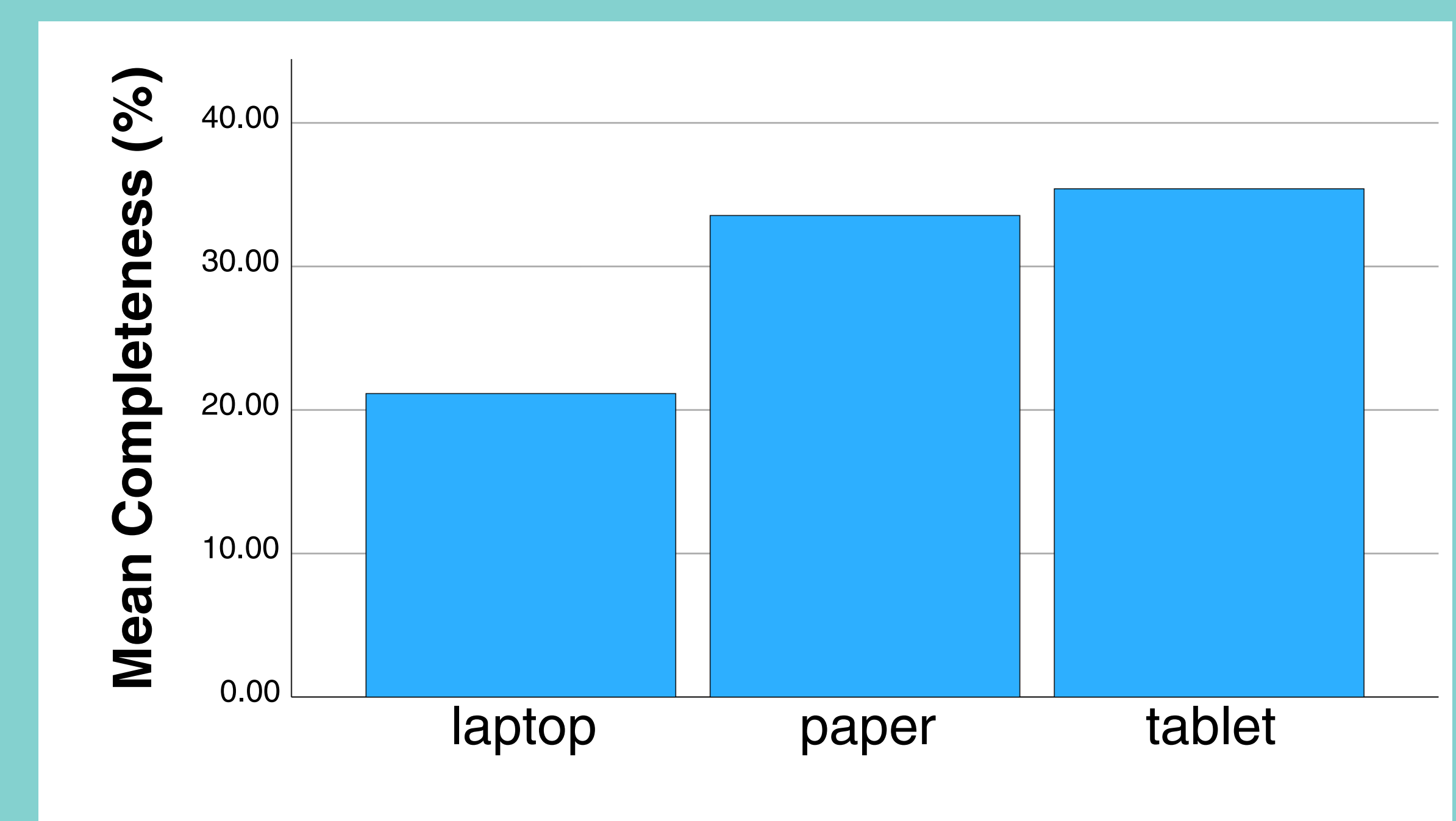
- Completeness:** for each iteration of the loop, a score was given when a value was written for a given variable. This was averaged across each labelled variable (Problem 1: total, count, val, and tracking the conditional).

## Completeness

### Problem 1

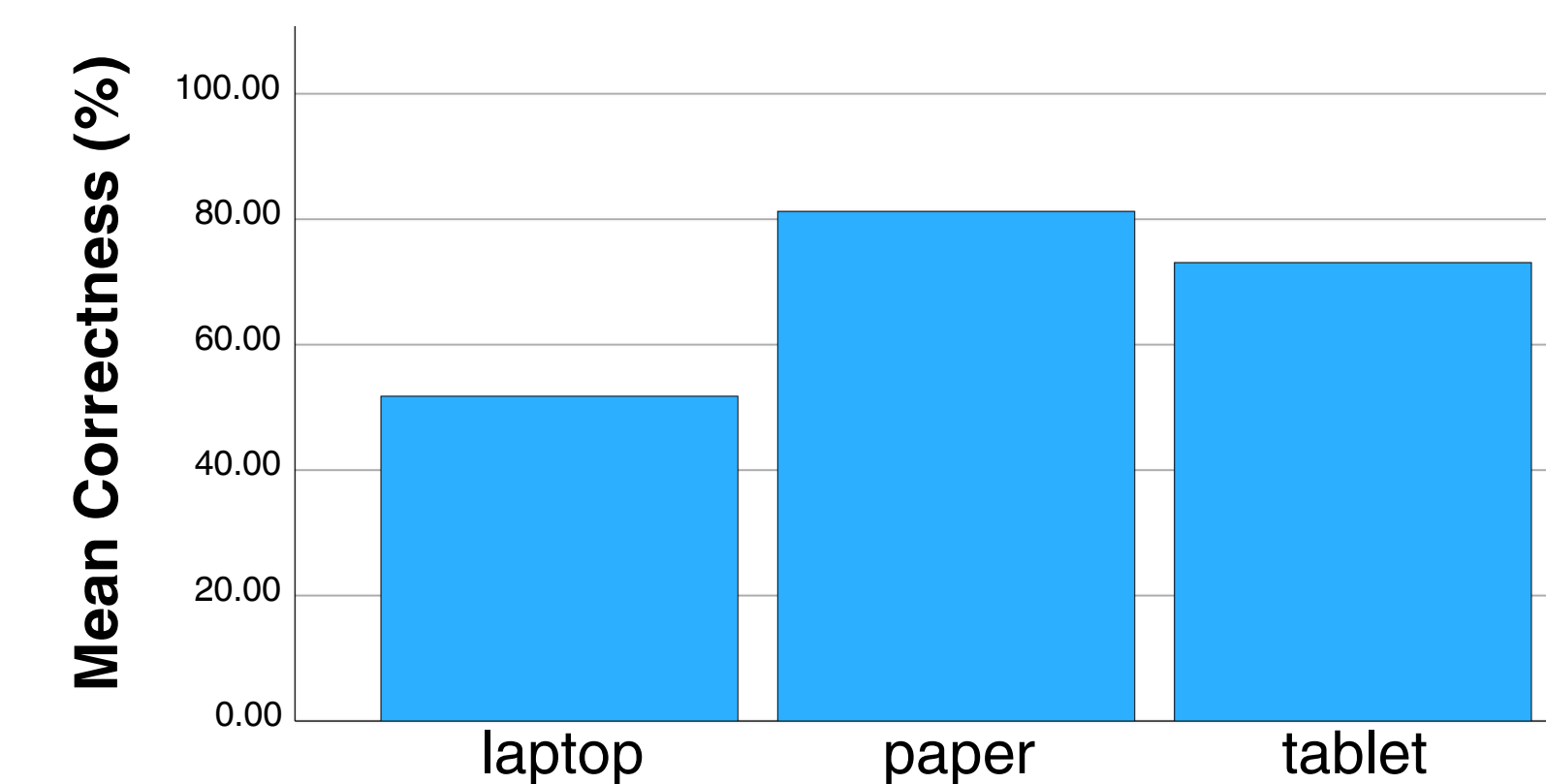


### Problem 4

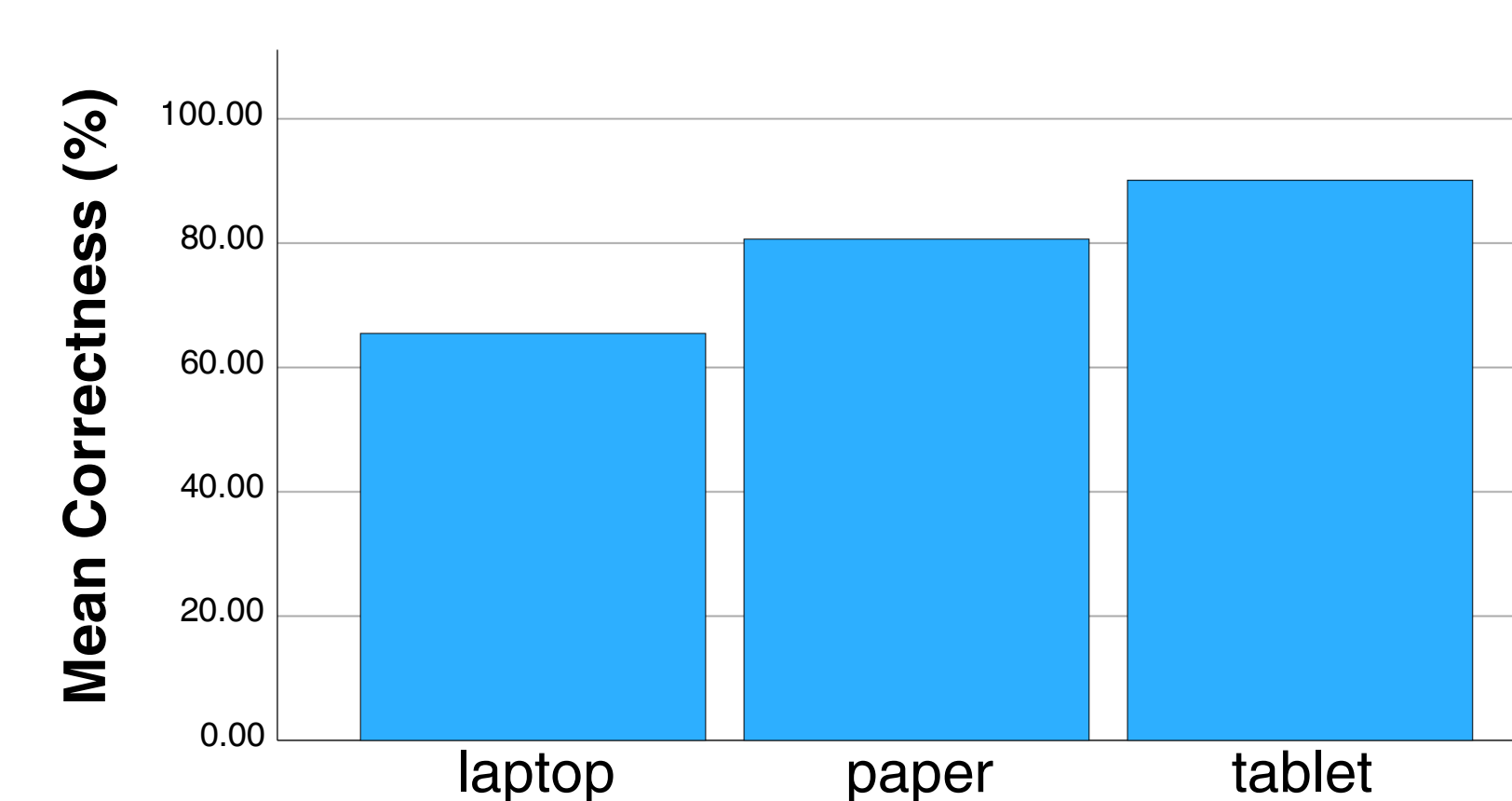


## Correctness

### Problem 1



### Problem 4

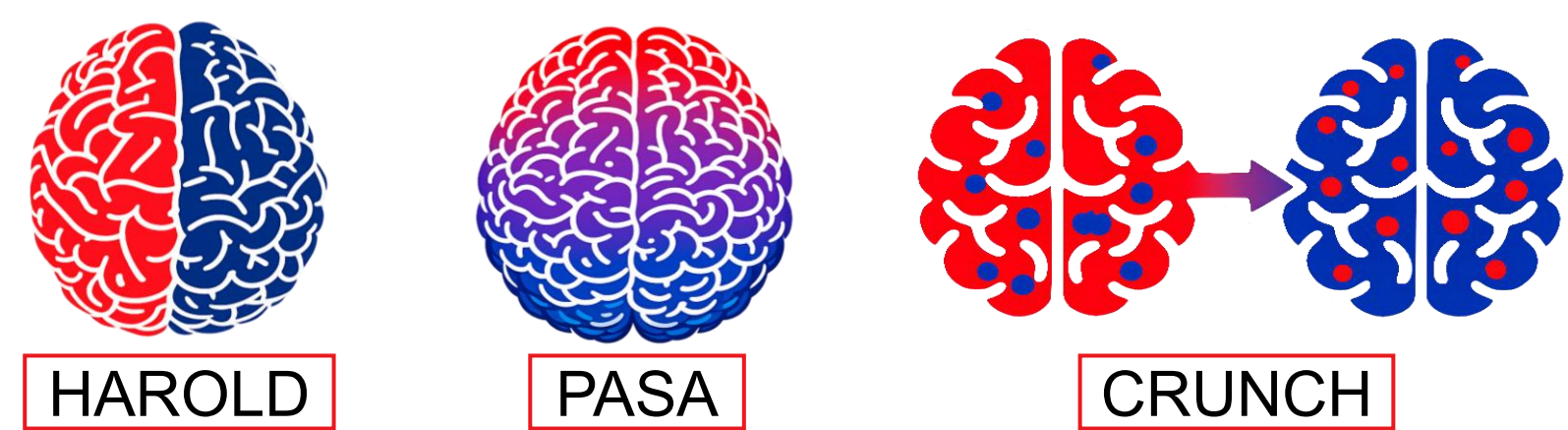


## 1 Background

- Some older individuals maintain cognitive performance despite age-related brain changes, suggesting compensatory neural mechanisms (Stern, 2002).
- Functional near-infrared spectroscopy (fNIRS) research shows cerebral oxygenation reflects both neural activity and cardiovascular health, which varies with aging and influences cognition (Suhr & Chelberg, 2013).
- However, how cardiovascular dynamics (e.g., heart rate variability; HRV) support cognitive reserve remains unclear. Integrating brain, behaviour, and physiology using accessible methods like fNIRS is critical for understanding cognitive resilience in aging.

## 2 Objectives

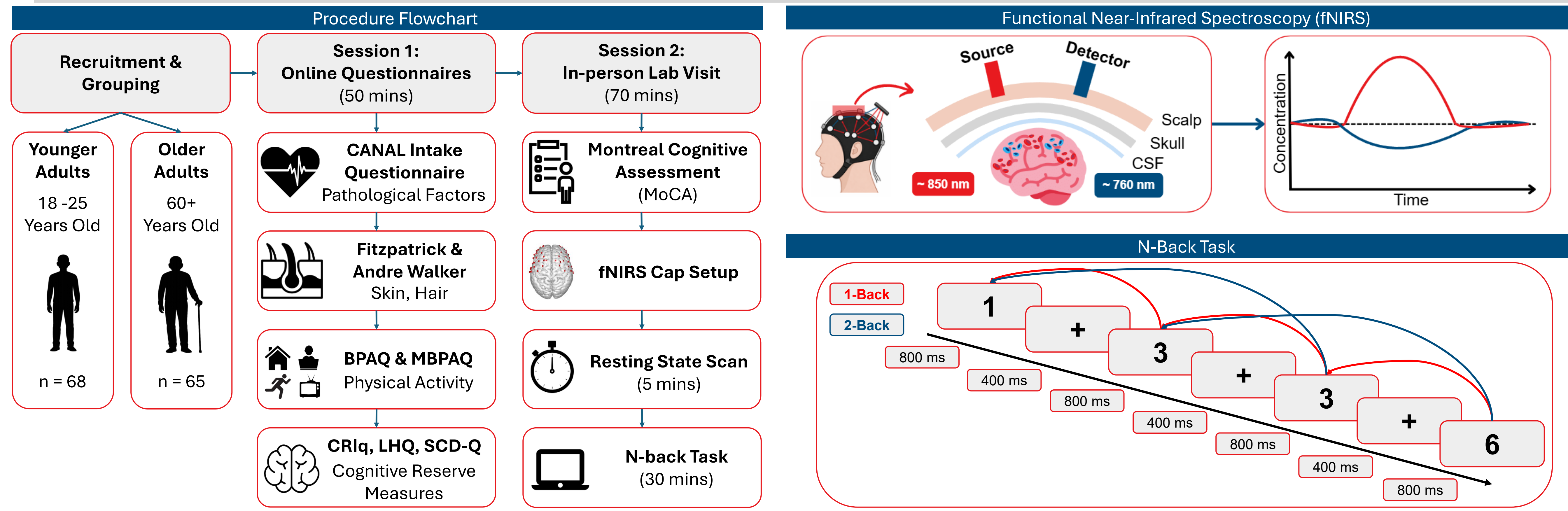
- Examine the relationship between cardiovascular health via physical activity, cognitive performance, and brain activation during an n-back task.
- Identify neural patterns supporting compensatory mechanisms.
- Evaluate fNIRS as a tool for studying the brain-heart interactions in the aging population.



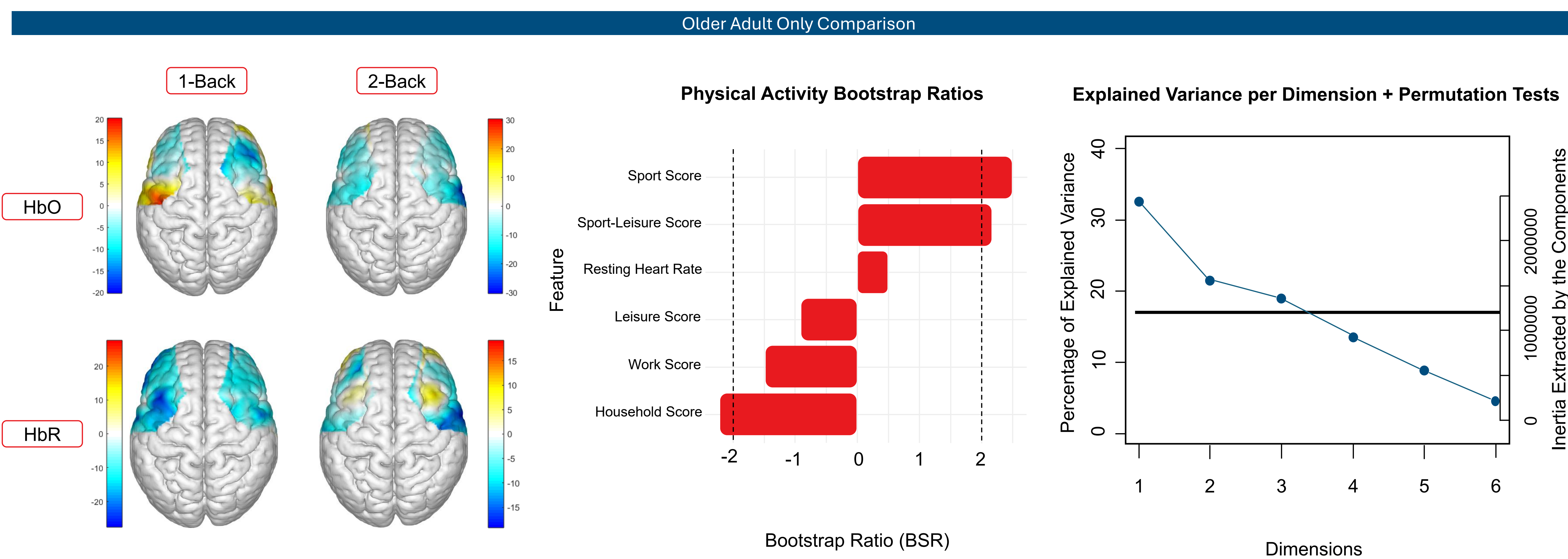
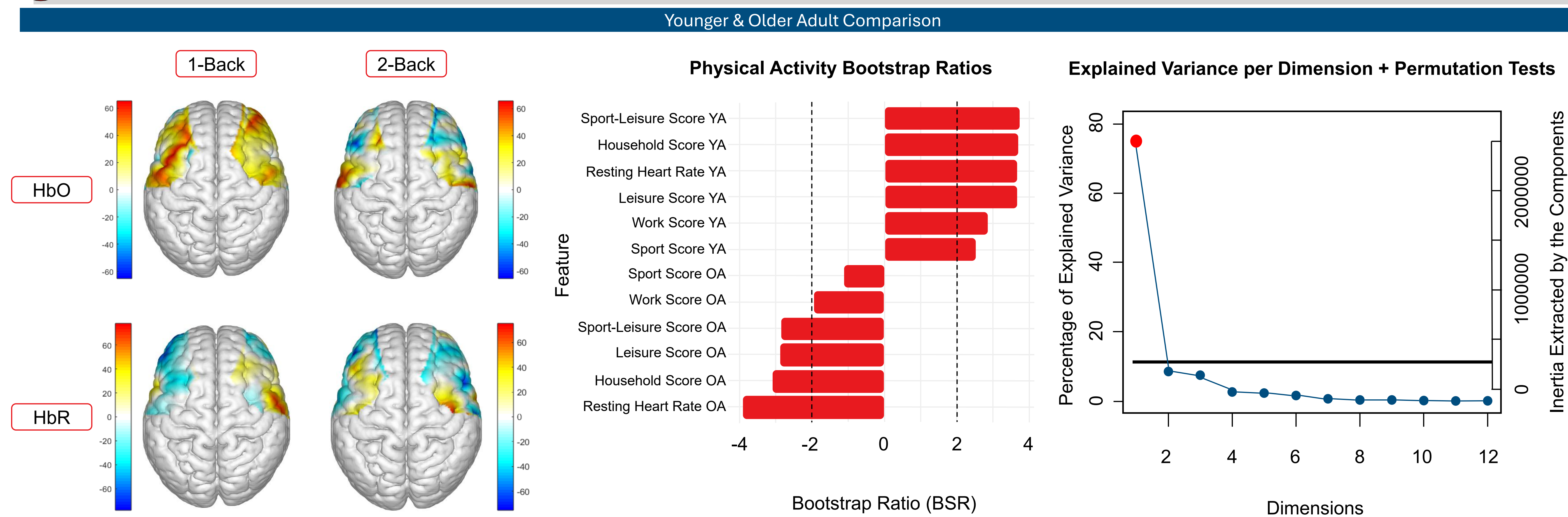
## 3 Hypotheses

- Based on research indicating a relationship between cardiovascular health, physical activity, cognitive reserve, as well as compensatory models of cognitive aging (HAROLD, PASA, and CRUNCH), the current study proposes 2 hypotheses:
- (H1)** Physically active older adults will outperform less active older adults on the n-back task.
- (H2)** Physically active older adults will have brain activation patterns consistent with younger adults.

## 4 Method



## 5 Preliminary Results



## 6 Conclusions

- These preliminary results provide partial, visual support H2 where physically active Oas may show brain activation patterns that are more like YAs.
- Across the YA-OA comparison, HbO showed the clearest task-related patterns, while HbR was more regionally variable, with greater differentiation emerging in the 2-back condition.
- Although the OA-only effects were not significant, the spatial trends suggest that higher sport and sport-leisure activities may support more youth-like neural recruitment during cognitive demand.

## 7 Future Directions

- Future work will build on the expected behavioural effects observed on the n-back task by testing whether better performance in physically active older adults is associated with more youth-like neural activation patterns.
- In addition, the present OA-only findings did not reach significance, which may reflect a small sample size rather than the absence of meaningful effect.
- A larger sample size will help determine whether the observed neural trends are reliable and whether physical activity is associated with both improved task performance and more youth-like brain activation in older adults.

## 8 References

- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8(3), 448–460. <https://doi.org/10.1017/s1355617702813248>
- Suhr, J. A., & Chelberg, M. B. (2013). Use of near-infrared spectroscopy as a measure of cerebrovascular health in aging adults. *Aging, Neuropsychology, and Cognition*, 20(2), 243–252. <https://doi.org/10.1080/13825585.2012.727976>

# Non-finiteness and Ergative Case Marking by Chester Leopold

## 1 Ergative and Accusative Languages

- (1) John saw *him*.  
 (2) *He* danced.  
 (3) \**Him* danced.  
 (4) \*Bill saw *he*.  
 (5) (Basque:)
- NOM/Accusative Alignment:  
 (Subject.NOM) transitive verb *object.ACC*  
 (Subject.NOM) intransitive verb

- a **Nekane-k** Miren eta Jon ikusi ditu.  
 Nekane-E Miren.A and Jon.A seen AUX.3pA.3sE  
 ‘Nekane saw Miren and Jon.’  
 b Miren eta Jon etorri dira.  
 Miren.A and Jon.A come AUX.3pA  
 ‘Miren and Jon came.’
- Ergative Alignment:  
*Subject.ERG* transitive verb (object.ABS)  
 (Subject.ABS) intransitive verb

What happens to the case marking of subjects in non-finite clauses?

In English, nominative case cannot appear on subjects in non-finite clauses (Rezac et al. 2014, Coon 2017):

- (7) John saw [*him* catch fish]. (accusative embedded subject)  
 (8) \*John saw [*he* catch fish]. (nominative embedded subject)

Bobaljik (1993) and Johns and Smallville (1999) claimed that ergative subjects cannot appear in non-finite environments. Nor can there be ergative agreement with the subject in these contexts:

- (9a) **Katu-ak** **sagu-ak** **harrapa-tzen** ikusi ditut. (Basque non-finite)  
 cat-d.pA mouse-d.pA catch-ing seen AUX.3pA.1sE  
 ‘I saw the cats catch the mice.’  
 (9b) **Katu-ek** **sagu-ak** **harrapa-tu** / **harrapa-tzen** dituzte-la (Basque finite)  
 cat-d.pE mouse-d.pA caught / catch-ing AUX.3pA.3pE-that  
 ikusi dut.  
 seen AUX.1sE  
 ‘I saw that the cats caught / were catching the mice.’

Some data from the following languages might be evidence for Bobaljik’s rule:

- Hindi (Bhatt 2005, Mahajan 2017)  
 Basque (Rezac et al. 2014, Tollan 2013)  
 Nukuoro (Drummond 2023)  
 Waripiri, though only partially (Laughren 2017)
- (10a) **mẽ=ne** [**ram=ko**/\***0** **kitab** **paɾ<sup>h</sup>-te** **hu-e**] **dek<sup>h</sup>-a** (Hindi non-finite)  
 I=ERG Ram =ACC/[NOM] book[NOM] read-IMP.F be-PF-OBI see-PF.M  
 ‘I saw [Ram read(ing) a book].’ (Event/proposition as direct object of ‘see’)  
 Perfective complement  
 (10b) **mẽ=ne** [**bari<sup>f</sup>-0** /**=ko** **paɾ-ti** **hu-i**] **dek<sup>h</sup>-i** (Hindi non-finite)  
 I=ERG rain. F. [NOM]/ACC fall-IMP.F be-PF-F. see-PF<sup>11</sup>  
 ‘I saw [rain falling].’ (Event as direct object of ‘see’)  
 (11) **Main** **maantaa-hu** **ki** [**Raam-ne** **ek** **kitaab** **paRhii**]  
 I believe that [Ram-Erg a book read] (Hindi finite)  
 ‘I believe that Ram read a book.’

Though how robust is Bobaljik’s rule given the many apparent exceptions?:

- a. Nominative subjects can appear in non-finite clauses in Western Portuguese (Legate 2006).  
 b. Legate (2006:238) notes that, “all cases are available in nonfinite (“subjunctive”) clauses” in Niuean, including ergative.  
 c. Mahajan (2017) cites a hindi sentence where an ergative subject might appear in a non-finite embedded clause.  
 d. Ergative agreement can appear in Ch’ol Mayan embedded “less-finite”/“non-finite” clauses (Coon & Vázquez Álvarez 2025.):

- (12) **Ngarrka-patu-rɬu** **ka-lu-jana** **puluku** **turnu-ma-ni,** (Waripiri)  
 man-PAUC-ERG PRESIMPF-3pl.SUBJ-3pl.OBJ bullock muster-NPAST  
 [karnta-patu-ku/karnta-patu-rɬu **miyi**/\***miyi-ku** **purra-nja-puru.**]  
 [woman-PAUC-DAT/woman-PAUC-ERG **food.ABS**/\***food-DAT** cook-NONFIN-TEMPC]  
 “The men are mustering cattle while the women are cooking the food.”  
 (My translation of the non-finite clause: ‘while the woman *cook* food’.)

- (13) **Kua** **kamata** [**ke** **hala** **he** **tama** **e** **akau**]  
 PERF begin [SBJV cut ERG child ABS tree] (Niuean)  
 ‘The child has begun to cut down the tree’ (M [21])

## 2 Are Hindi raising constructions an exception to Bobaljik’s rule?

- (14) **Kabir-ne** **bahut** **baɾii** **galtii** **kar** **d-i** (\***thii**)  
 Kabir(masc)-erg very big mistake(fem) do give-perf.fem be.pst.fem  
 lag-tii thii  
 seem-imperf.fem be.pst.fem  
 ‘Kabir seemed to have made a big mistake.’  
 (“somewhat marginal for some Hindi speakers”)

Mahajan (2017) claims that no matter what, the subject “Kabir” must get its ergative case as the transitive subject of “to have made” inside of the non-finite clause.

This satisfies the Hindi rule that ergative case can only be assigned in perfective clauses.

Mahajan considers two possible structure for (14):

- i. The ergatively marked “Kabir” stays inside of the non-finite clause and breaks Bobaljik’s generalization.  
 ii. “Kabir” moves with its ergative case to the subject of the *finite* verb “seemed” through raising out of the non-finite clause:

- (I) [ seemed [Kabir.Erg to have made.Perf a big mistake]]

Finite non-finite

- (II) [Kabir. Erg seemed [Kabir.Erg to have made.Perf a big mistake]]

## 3 Basque “Raising to Ergative”

Sentence (14) looks similar to Rezac et al.’s (2014) Basque “Raising to Ergative” constructions.

These Basque sentences contain an infinitive verb and the modal verb “behar”/“must”/“needs to”:

- (15) **Jon-ek** **eta** **Miren-ek**<sub>i</sub> **etorri** **behar** **du-te**<sub>i</sub>.  
 Jon-E and Miren-E come must AUX-3pE.  
 ‘Jon and Miren must come.’

- (16)  $(S_{ERG}) T_{ERG}$  **behar** [ $V_{ABS}$  ... ( $S_{ABS}$ ) ...] (Raising to ergative)  
 (iii) T-Agree (Case if S Moves to Spec,T) (ii) v-Agree/Case (v) Move/EPP (optional)

- (17) [ $(S_{ABS}) T$  [ $V_{ABS}=S$  ( $S_{ABS}$ ) ]]
- (Regular unaccusative)

Though there are interesting reasons to question that the Hindi (14) and the Basque (15) are on a par.

## 4 Bhatt et al. (2011) on Hindi Raising

Bhatt et al. (2011:5) claim that: “There is no straightforward equivalent to English-style raising constructions in Urdu/Hindi”.

Bhatt et al. write that, “*lag* ‘attach to’ cannot be used with *non-finite* complements as the English ‘seem’ (e.g. *John seems to be going home.*)” (6).

Therefore, Bhatt et al. might not assume with Mahajan that (14) can be well-formed for certain speakers.

Bhatt et al. nonetheless identify interesting modal + infinitive constructions in Hindi.

Bhatt et al. claim that these constructions cannot have ergative case even when the non-finite verb is transitive.

Bhatt et al. conclude that these constructions are best analyzed as involving control:

- (18) **yasin=ne**/\***ko** **ye** **ki-ya** (Hindi regular transitive perfective)  
 Yasin.M.Sg=Erg/Dat this.Nom do-Perf.M.Sg  
 ‘Yasin did this.’

- (19) **yasin=ko** **ye** **kar-na** **cahiye** (Hindi modal + infinitive)  
 Yasin.M.Sg=Dat this.Nom do-Inf.M.Sg need.Sg  
 ‘Yasin needs to do this.’

(Though I don’t think (19) is in the perfective)

- (20) **Yassin** **needs** [**PRO** to do this]. (control structure)

## 4 What is non-finiteness?

Theorists typically define non-finiteness as the absence of tense and agreement in a clause and the inability for a non-finite verb to appear as the main verb of clause (Santorini and Kroch 2007, Carnie 2021):

- (21) a. I want [him to eat asparagus]. (non-finite)  
 b. \*I want [him to eats asparagus]. (finite)

- (22) a. He gives both of the charities money. (finite)  
 b. \*He give both of the charities money. (non-finite)

However, Johns and Smallville (1999) note that linguists do not consider all languages to behave the same with respect to non-finiteness, (this makes it challenging to correctly categorize the sentences that are relevant to testing Bobaljik’s hypothesis):

Main clause	Tense	Agreement	
+MC	−T	−A	Russian Infinitivals; Middle Welsh infinitivals
−MC	+T	−A	Tamil participles; Lezgian participles
−MC	−T	+A	European Portuguese infinitivals; Dravidian particples; Welsh infinitivals; Turkic infinitivals
−MC	−T	−A	English infinitivals; Chinese “non-finite”

Johns and Smallville note that linguists tend to classify non-finiteness in a given language when a verb category or clause exhibits at least two of the above three features.

For example, unlike English, Middle Welsh infinitives can appear as the main verb in the sentence:

- (23) **a’e** **dyuot** **ynteu** **yr** **llys** (Middle Welsh)  
 and-3ms **come.inf** he to-the court  
 ‘And he came to the court.’ (Tallerman 1998b, citing Evans 1989)

## 5 Legate’s (2006) Absolutive Hypothesis

The idea that case assignment is deficient in non-finite environments is discussed elsewhere with respect to absolutive case (Legate 2006; Coon, Mateo Pedro & Preminger 2014; Polinsky 2017; Coon & Vázquez Álvarez 2025):

Legate proposed that in one class of ergative languages– “absolutive as nominative”, absolutive case cannot be assigned in non-finite environments to intransitive subjects and to transitive objects.

- (24)

This is because the functional head that assigns absolutive case in these languages “T” is deficient unlike the “T” in finite clauses.

In Legate’s second class of languages– “absolutive as default”, absolutive case cannot be assigned to intransitive subjects in non-finite environments. However absolutive case can be assigned to transitive objects in non-finite environments.

- (25)

This is because a functional head in the structure other than “T” can assign absolutive case to the object, namely ‘little v transitive’.

## 6 Outstanding Questions:

Are parameters relevant for describing the difference in case assignment across Legate’s two language groups?

Universal Grammar must permit children to learn ergative *and* accusative languages along with their rules for case assignment in non-finite environments.

One wants to know how theories of parameters may be relevant to this discussion.



# Semantic Association Transfer in Bilinguals (a Proposal)



Fengwei Liu, Olessia Jouravlev

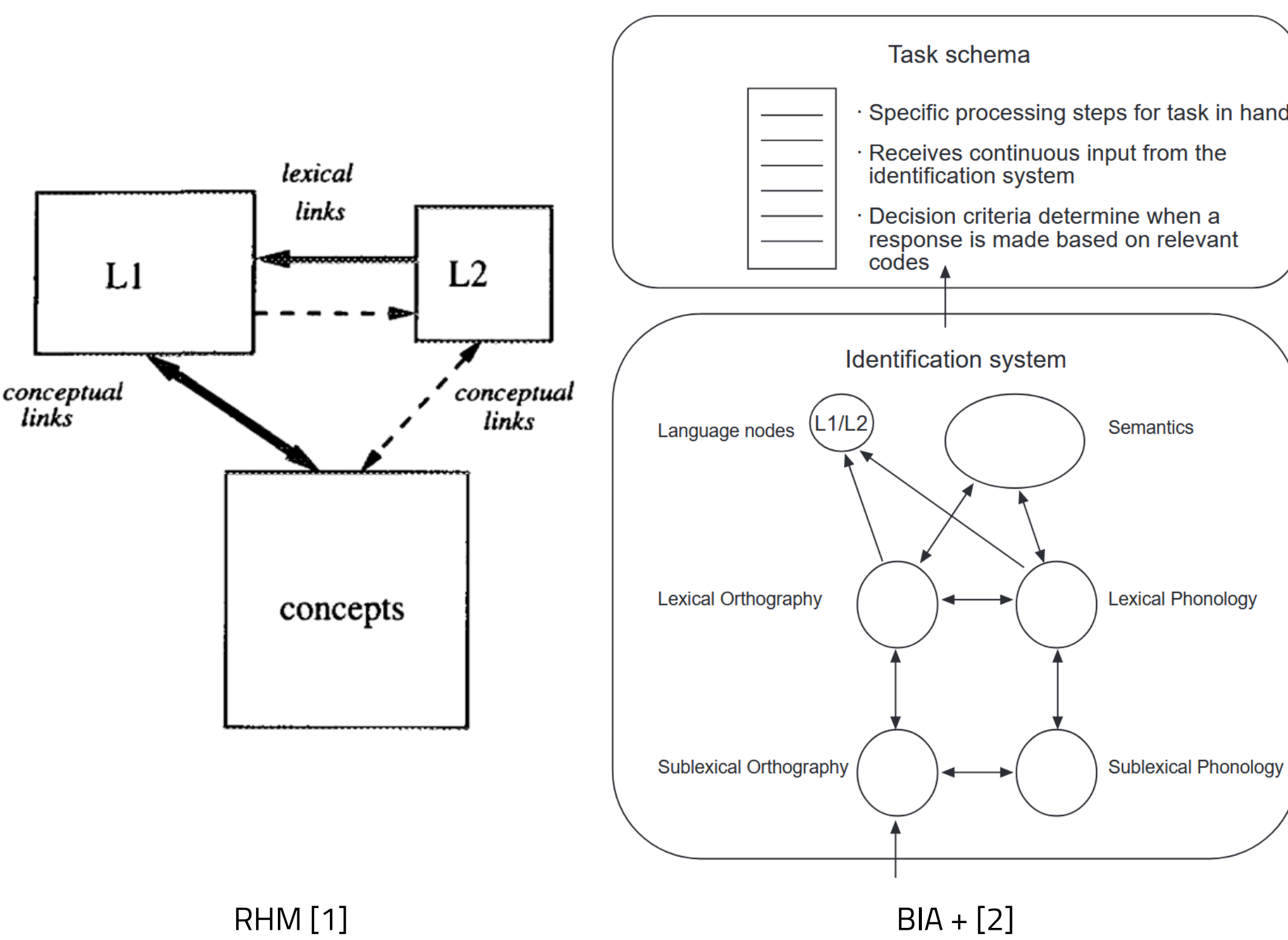
Linguistic Neurodiversity Lab, Department of Cognitive Science, Carleton University

## BILINGUAL SEMANTICS

How are the two linguistic systems organized in the bilingual brain?

An integrated, interactive structure

- Language-invariant representations
- Connections across linguistic levels



Language-agnostic lexical access

- Translation equivalents prime semantic activation [3]
- Bilingual patterns for categorizing daily objects converge across languages [4]
- A shared brain network for L1 and L2 lexico-semantic processing [5]

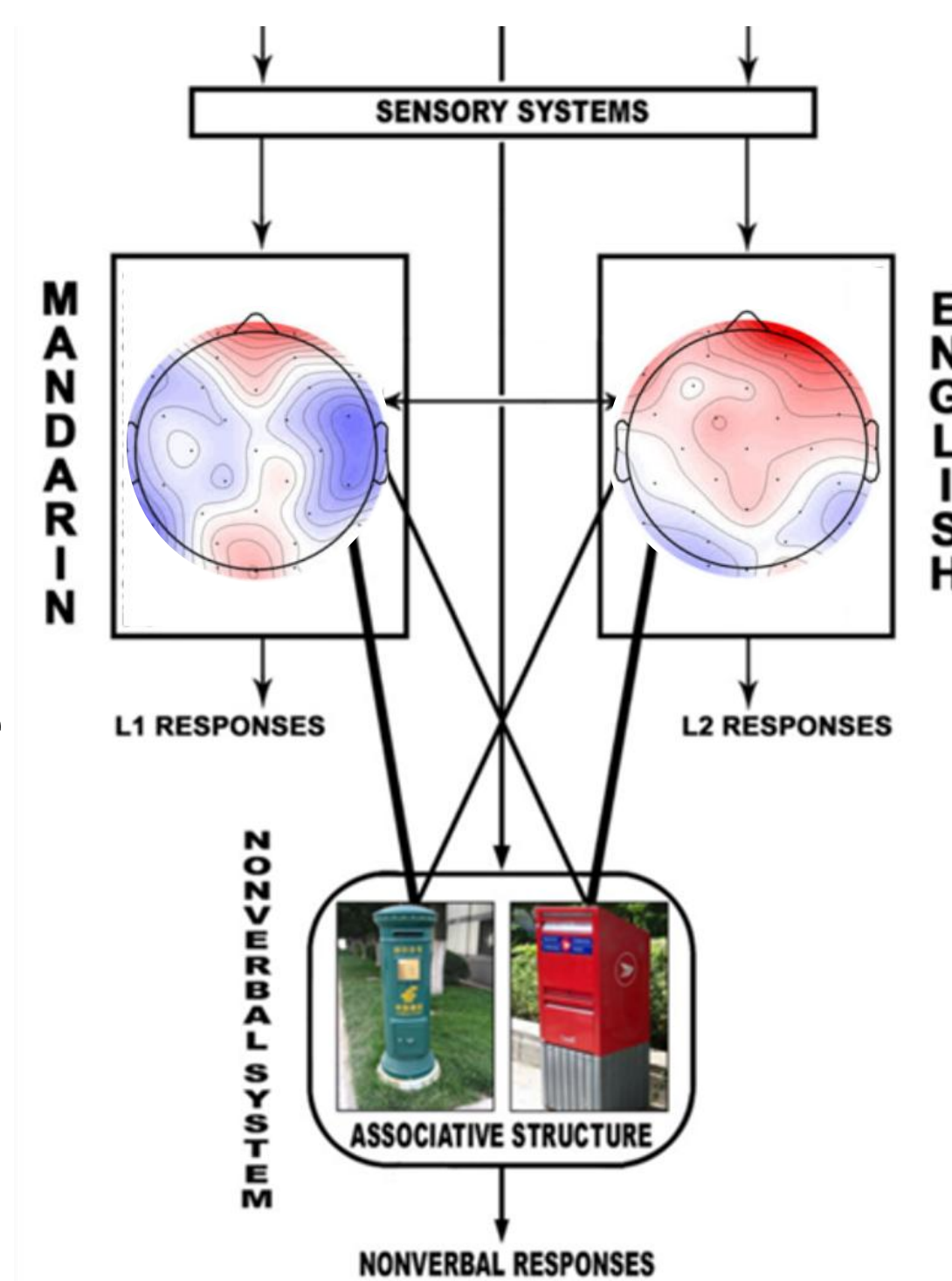
Translation equivalents are not conceptual equivalents.

A translation pair may have slightly different referents

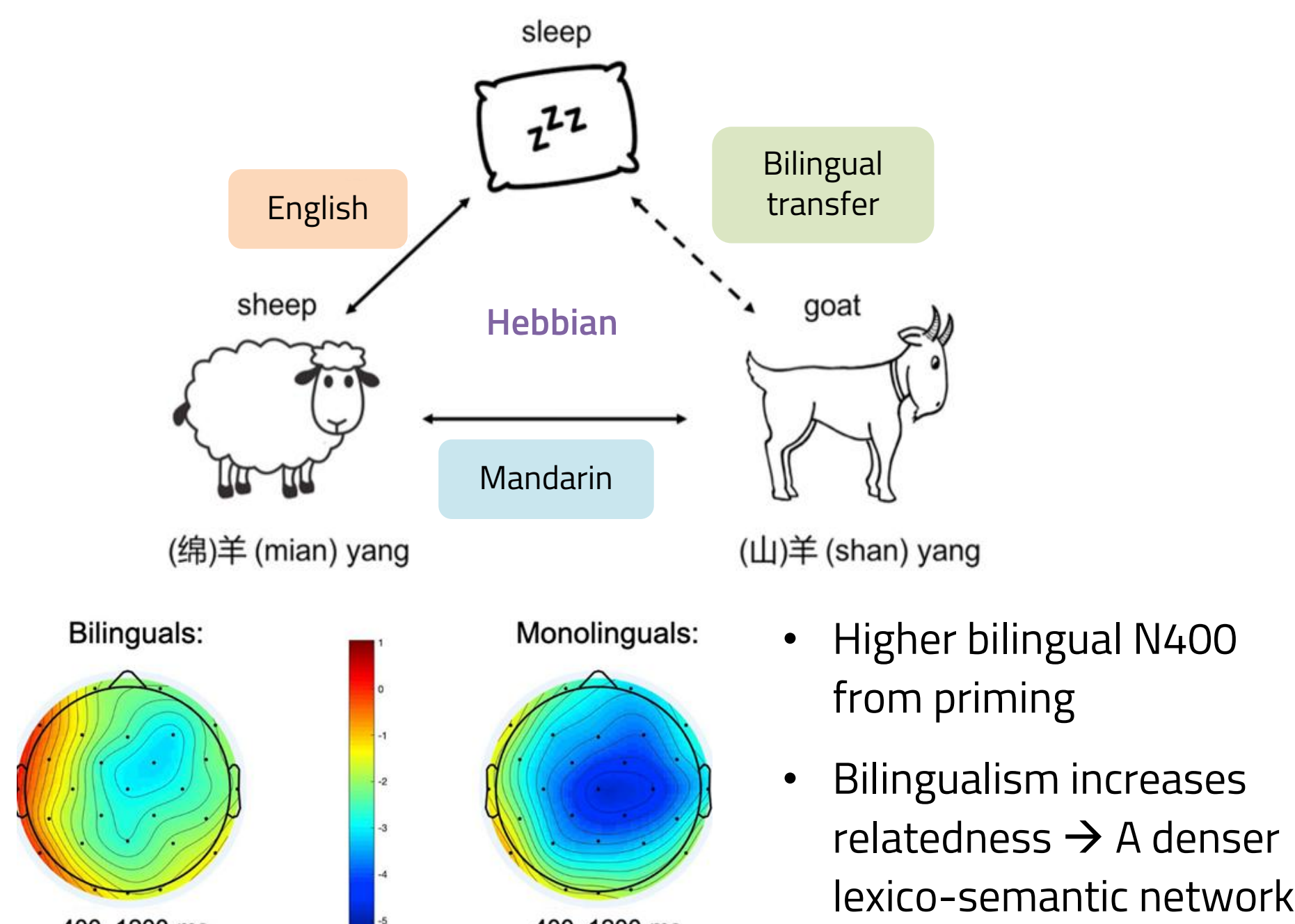
frustration ≠ переживать

Conceptual representations diverge between culture-specific associations [6]

- Interaction effect of language and culture on ERP and RT
- Culturally congruent imagery activate corresponding language
- Language-specific conceptual representations exist for translation equivalents



Associative structures transfer across languages [7]



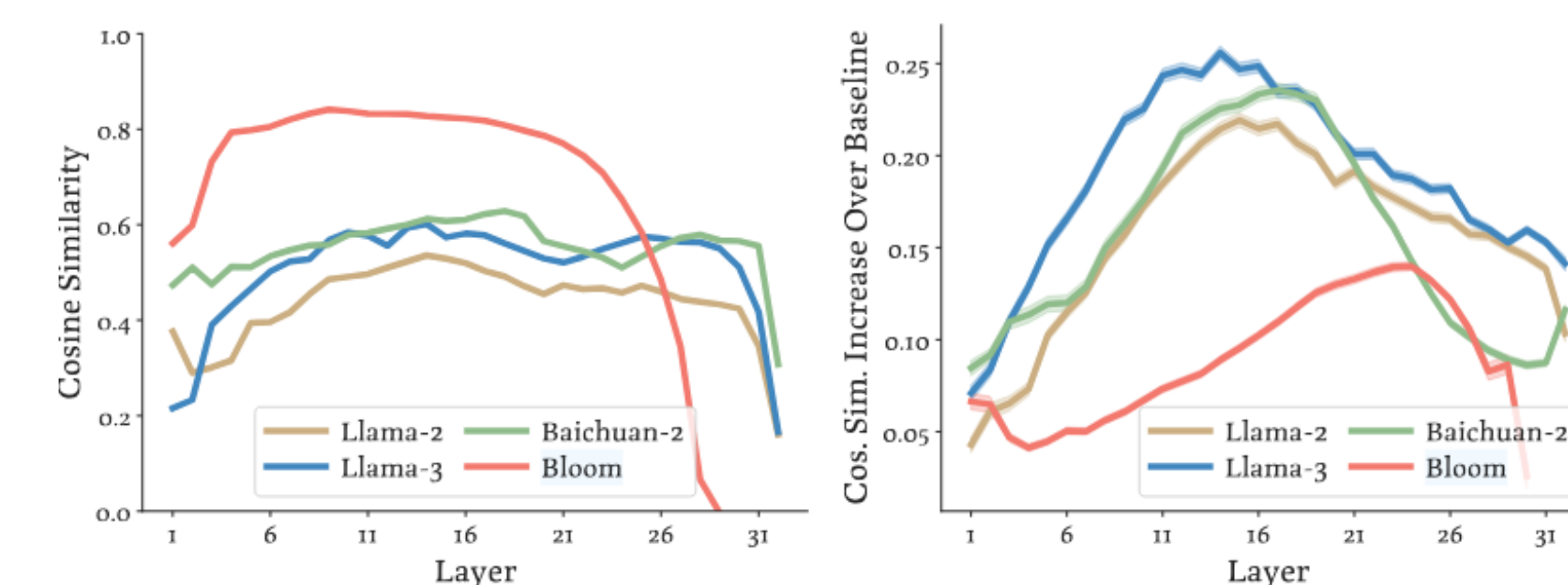
Are semantic associations specific to one language transferred into the other?

## MULTILINGUAL LLMs

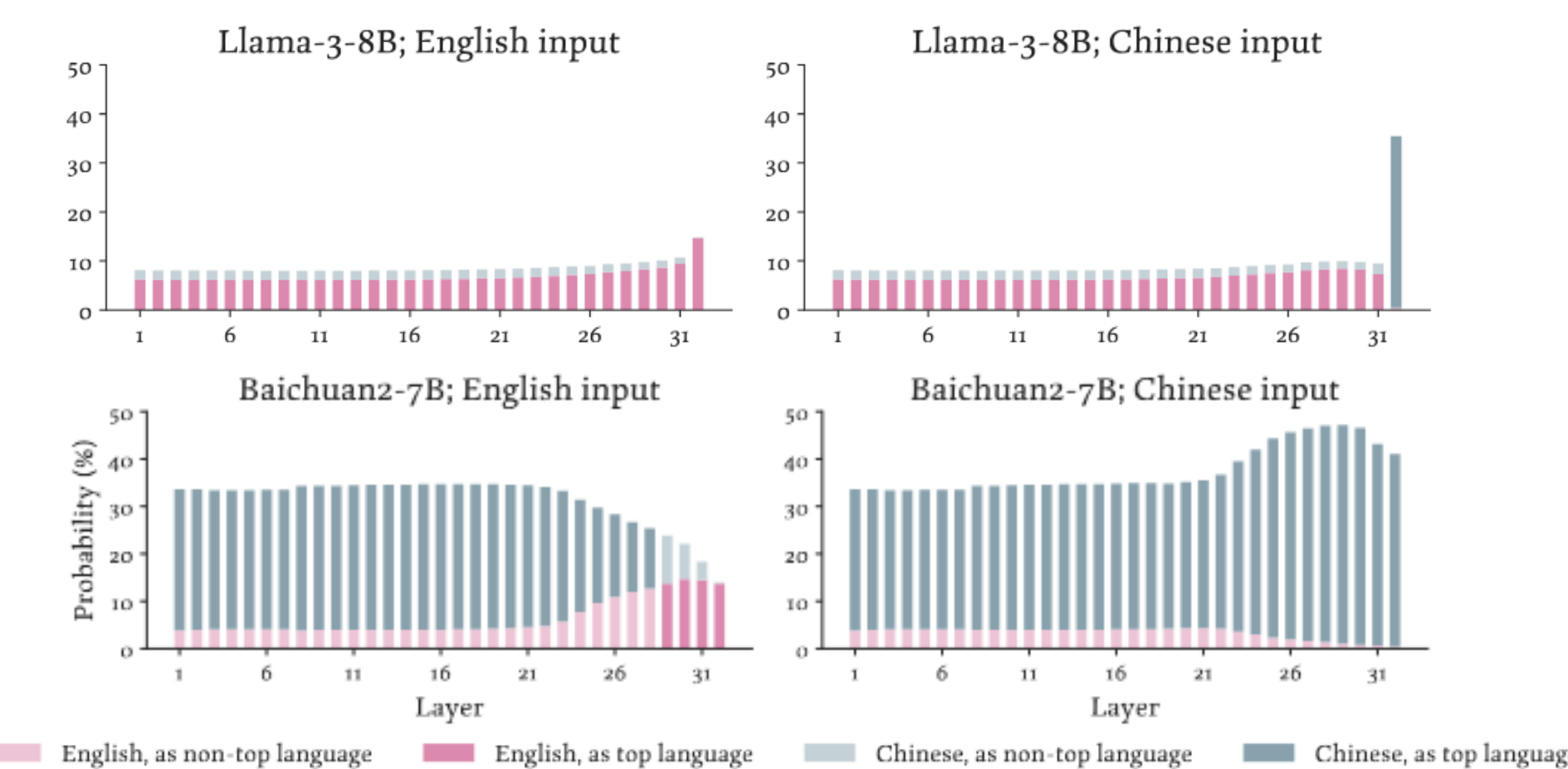
LLMs have super-human multilingual abilities. Are languages organized in LLMs in similar ways as the human brain?

The Semantic Hub Hypothesis [8]

- Cross-lingual & cross-modal representations in the middle layers
- Similar representations between translated sentences



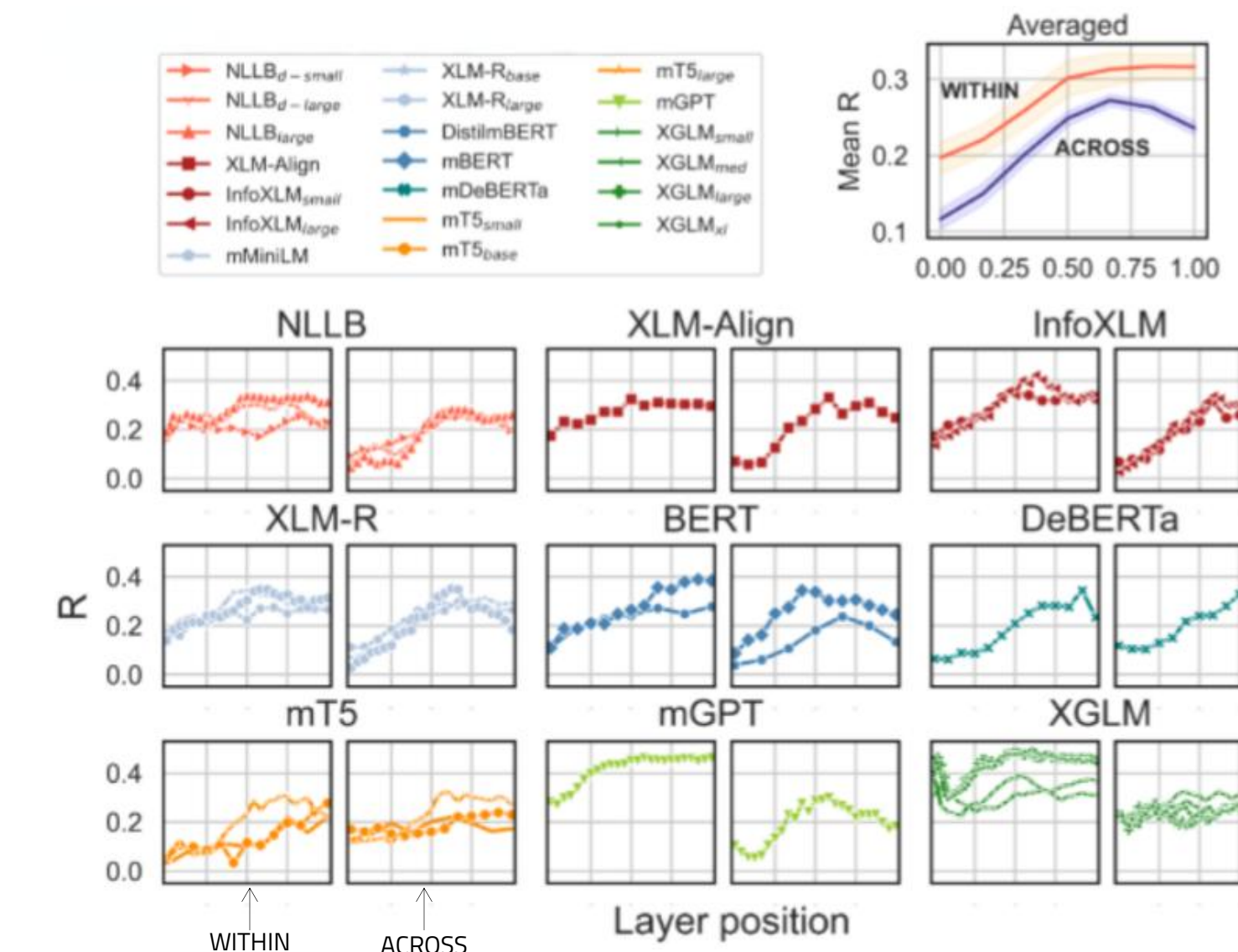
Early and middle layers are aligned with the dominant language - logit lens



Cross-lingual interventions steer LLM output sentiment - ActAdd

Text Lang.	Steering Dir.	Steering Lang.	Sentiment	Disfluency (↓)	Relevance (↑)
Spanish	↓	None	0.143±0.022	7.35±1.19	0.861±0.002
		Spanish	0.125±0.034	10.54±2.39	0.842±0.004
	English	0.139±0.026	8.75±2.20	0.857±0.002	
	↑	Spanish	0.175±0.035	7.98±2.04	0.856±0.002
		English	0.159±0.026	7.35±1.01	0.859±0.003

LLM embeddings predict BOLD response across languages [9]



No study on cross-linguistic conceptual transfer in LLMs

## RESEARCH QUESTIONS

RQ 1: Do language-specific semantic associations carry over to the other language in bilinguals?

H: Compared to monolinguals, bilinguals exhibit smaller N400 effect and faster RT with words connected only in the other language.

RQ 2: To what extent do multilingual Transformer-based LLMs show similar patterns of cross-lingual association transfer on a representational level?

H: Conceptual organization across languages converge in multilingual LLMs. Balanced training data results in larger degrees of convergence.

RQ 3: What representational structure contributes to conceptual convergence in LLMs?

H: Language-agnostic neurons in the middle layers of multilingual LLMs output similar encodings of these concept pairs.

## EEG SEMANTIC PRIMING

	M+ E+	M+ E-	M- E+	M- E-
M-E Bilingual	N400 --	N400 -	N400 --	N400 +
Monolingual	N400 --	N400 +	N400 --	N400 +

Participants: 20 Mandarin-English bilinguals and 20 English monolinguals

- Balanced early bilinguals
- Minimal second-lang proficiency
- LEAP-Q & LexTALE for screening
- AoA & AOP

Stimuli: 40 prime words in 4 associative structures each

- Association norms: Small World of Words (10,024 M & 12,217 E cues)
- Step 1: translation equivalent norms for 50,000+ Mandarin words [10]
- Step 2: association pairs that received the most diverging percentages of response in the 2 languages
- 4 counterbalanced lists with all primes once in each session

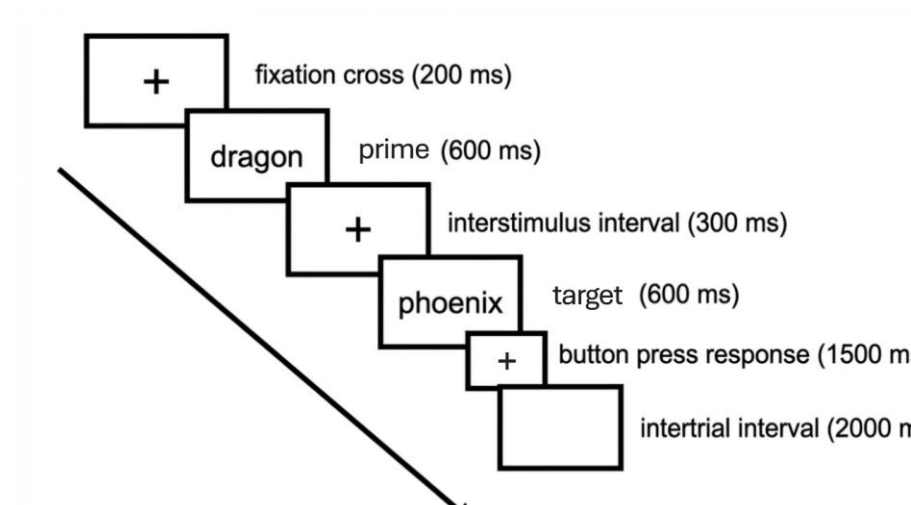
	E+	E-		E+	E-
dragon			line		
M+	fantasy	phoenix	M+	queue	army
M-	scales	blue	M-	dance	pigeon

- Confounds:
- word frequency
  - concreteness
  - cognates
  - co-occurrence

Paradigm: priming and semantic relatedness task

ERP component

- Epoch: -200 to 800 ms
- N400: 300-500 ms
- Centro-parietal electrodes



## PROBING LLMs

Models:

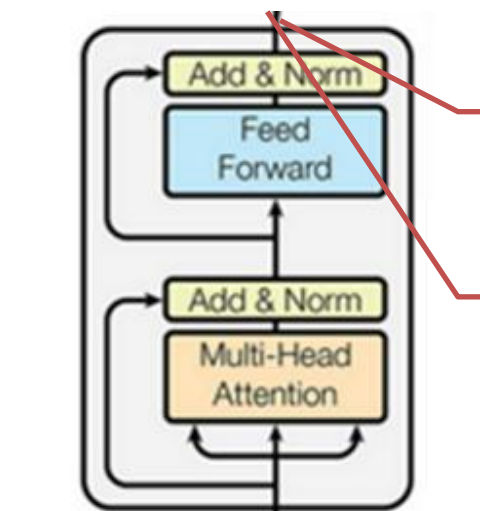
- Monolingual: Goldfish\* (125M, autoregressive decoder-only)
- Bilingual: GLM-130B (bidirectional encoder-decoder)
- Multilingual: GPT2\* (124M, autoregressive decoder-only), GPT2 large (774M), BLOOM (176B, decoder-only, balanced training data)

The same set of stimuli will be used to test if bi-/multilingual LLMs resemble human-like semantic transfer.

Prompting: "Rate the conceptual relatedness between these two words from 1 to 10: [prime], [target]. Make sure you only return a number and nothing more. Rating:"

Representational Pooling to decontextualize word embeddings:

- For each word, sample 100 sentences from natural corpora. Senses of polysemous words will not be controlled to simulate the experiment.



- Average token vectors to extract layer-wise word representations
- Pool word representations across 100 contexts using the average

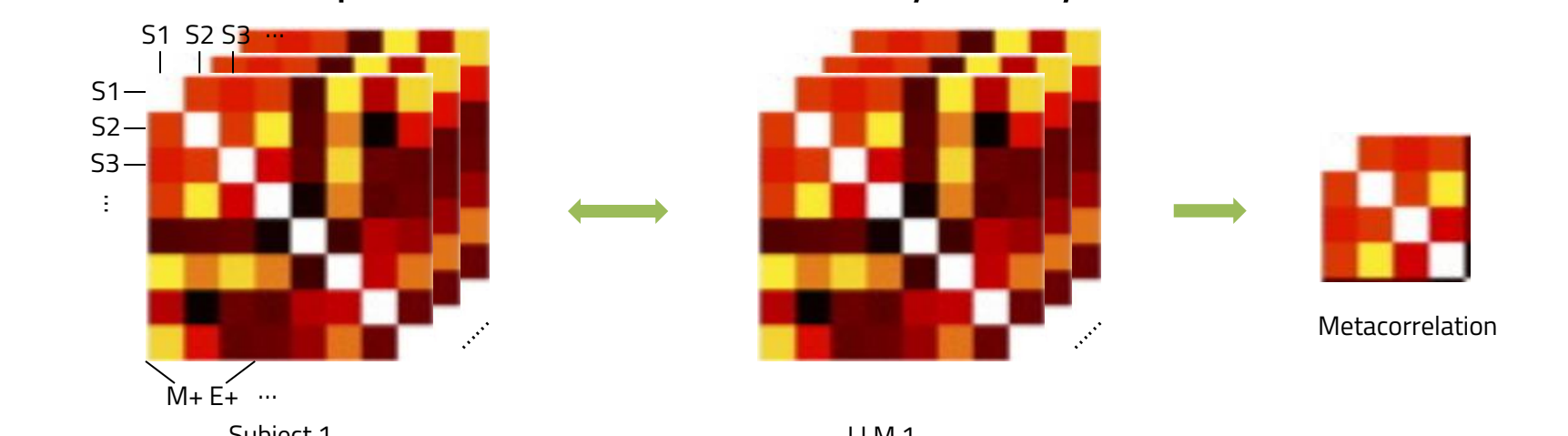
- Both EN and CN word representations will be extracted
- Sparse Autoencoders to interpret the concepts encoded by each word.

## ANALYSIS

RQ 1 – ERP & behavioural data: mixed-effect GLM

- Fixed effect: Associative structure, Bilingualism, Electrodes
- Random effect: Participant

RQ 2 – Representational Similarity Analysis



RQ 3 – Layer-wise cross-lingual alignment

- Cosine-similarity
- Retrieval precision in bilingual dictionary induction (precision@k) with Cross-domain Local Scaling after concept space alignment [11]
- Sparse autoencoders for interpretability

## REFERENCES

- Kroll JF, Stewart E. Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*. 1994 Apr 1;33(2):149–74.
- Dijkstra T, Heuven WJB van. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*. 2002 Dec;5(3):175–97.
- Altarriba J, Basnight-Brown DM. Methodological considerations in performing semantic- and translation-priming experiments across languages. *Behav Res Methods*. 2007 Feb;39(1):1–18.
- Armeel E, Storms G, Malt BC, Sloman SA. How bilinguals solve the naming problem. *Journal of Memory and Language*. 2005 Jul 1;53(1):60–80.
- Sulpizio S, Del Maschio N, Fedeli D, Abutalebi J. Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*. 2020 Jan 1;108:834–53.
- Pan X, Xiong A, Jouravlev O, Jared D. Conceptual representations in bilinguals: An ERP approach. *Bilingualism: Language and Cognition*. 2021 Nov;24(5):932–44.
- Ning S, Hayakawa S, Bartolotti J, Marian V. On language and thought: Bilingual experience influences semantic associations. *Journal of Neurolinguistics*. 2020 Nov 1;56:100932.
- Varda AG de, Malik-Moraleda S, Tuckute G, Fedorenko E. Multilingual Computational Models Reveal Shared Brain Responses to 21 Languages [Internet]. *bioRxiv*; 2025 [cited 2026 Feb 20]. p. 2025.02.01.636044.
- Wu X, Yu XV, Yogatama D, Lu J, Kim Y. The Semantic Hub Hypothesis: Language Models Share Semantic Representations Across Languages and Modalities [Internet]. *arXiv*; 2025 [cited 2026 Feb 19].
- Liu Y, Chen S, Yang Y. Semantic alignment: A measure to quantify the degree of semantic equivalence for English-Chinese translation equivalents based on distributional semantics. *Behav Res*. 2025 Jan 8;57(1):51.
- Peng Q, Søgaard A. Concept Space Alignment in Multilingual LLMs. In: Al-Onaizan Y, Bansal M, Chen YN, editors. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* [Internet]. Miami, Florida, USA: Association for Computational Linguistics; 2024 [cited 2026 Feb 16]. p. 5511–26.



# Behavioural Signatures of Dopaminergic Variation: A Machine Learning Approach to N-back Performance



Hannah Matresky, Jaiden Bachelder, Mazzy Beasley, Natalie Tutu

CGSC 3601: AI and Cognitive Science, Department of Cognitive Science, Carleton University, Ottawa, Ontario, Canada

## Introduction

- COMT genotype influences executive functioning, particularly working memory and response inhibition (PFC)
- The COMT Val158Met polymorphism affects dopamine levels by altering enzyme activity<sup>1</sup>:
  - Val/Val: higher COMT activity → lower dopamine → poorer working memory performance
  - Met/Met: lower COMT activity → higher dopamine → better working memory performance
- Traditional analyses may miss nonlinear relationships (e.g., dopamine's inverted-U effect)
- Can working memory performance (n-back accuracy) predict underlying COMT genotype?

## Hypothesis

- We predict higher N-back accuracy scores in Met/Met individuals and lower scores in Val/Val individuals

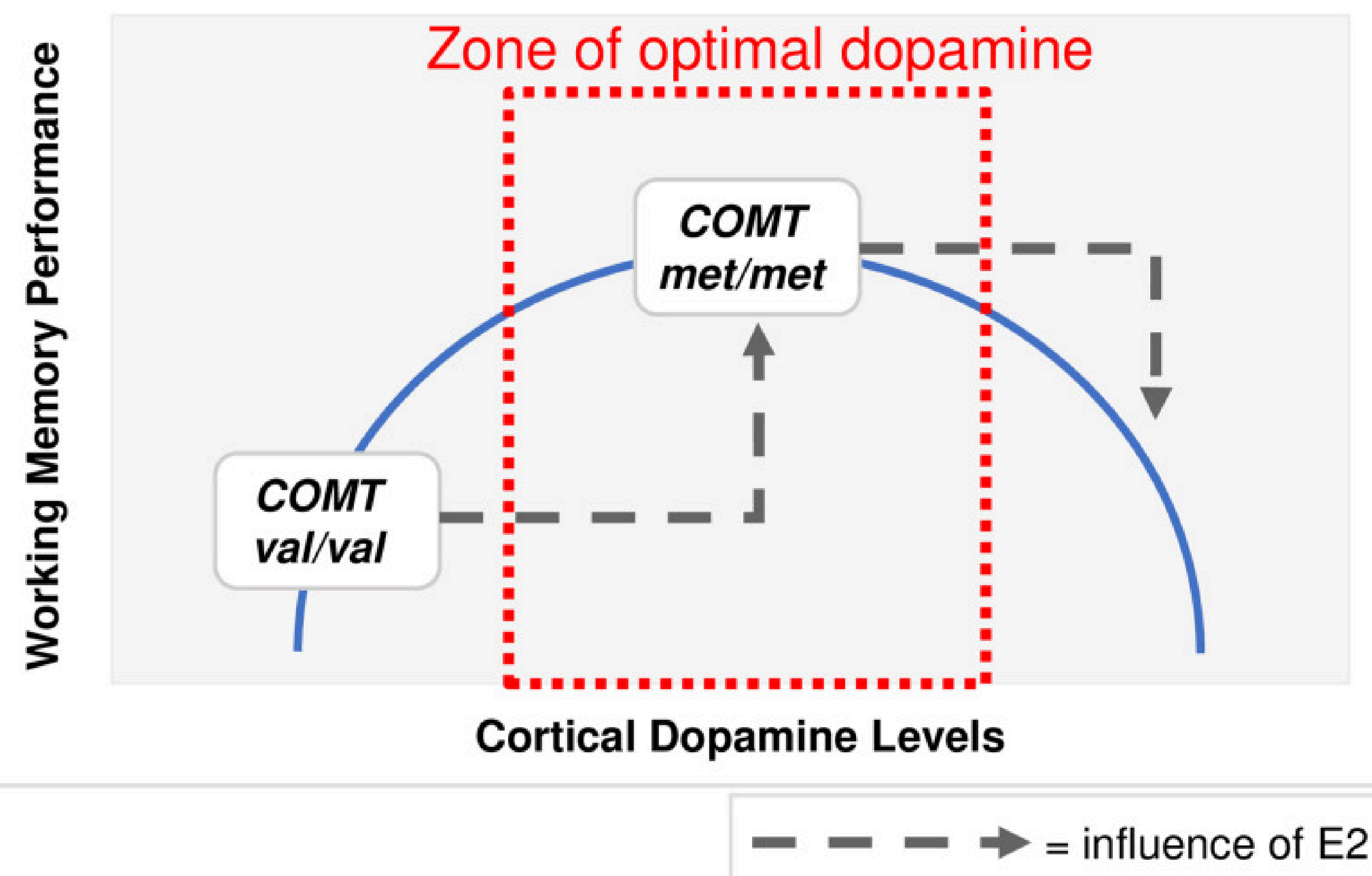


Figure 1. The inverted-U dopamine response curve and working memory in the Prefrontal Cortex (PFC)<sup>1</sup>

## Methods

### Dataset & Participants

- Sample: 119 young women (ages 18-30) with both behavioural + genotype data<sup>2</sup>
- Measures: n-back task performance + COMT Val158Met genotype (Val/Val, Val/Met (Het), Met/Met)
- Created participant-level features:
  - Mean RT (1-back, 2-back)
  - Total hits (correct responses)
  - Total false alarms (incorrect responses)
  - Genotype classification (Val, Het, Met)
- Final dataset: 119 participants × 8 features

### Models implemented in WEKA:

- ZeroR (baseline)
- Naïve Bayes (probabilistic)
- J48 decision tree (rule-based)
- Random Forest (ensemble, nonlinear)

### Evaluation Metrics

- Accuracy: Overall correct classifications
- Sensitivity (TPR): Correct identification of each genotype class
- Specificity (TNR): Correct rejection of non-target classes

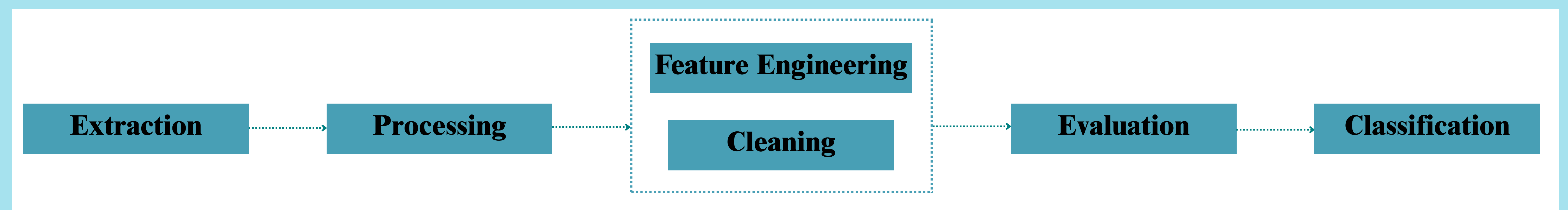


Figure 2. Overview of the data processing and machine learning pipeline.

## Results

- ZeroR (baseline): 47.88% accuracy; always predicted Het (majority class)
- Naïve Bayes: 27.73% accuracy (below baseline); overpredicted Met, highest specificity
  - Achieved higher specificity (TNR) than other models
- J48: 40.34% accuracy; no meaningful splits → defaulted to Het
- Random Forest: 40.34% accuracy; similar to J48, no improvement with complexity

## Conclusion

- No model outperformed the baseline (ZeroR)
- Strong tendency to overclassify the majority class (Het)
- Minority classes (Val, Met) poorly identified across models
- Behavioural features showed limited ability to distinguish genotypes

## Acknowledgments

Thank you to our Professor Dr. Mary Kelly and our Teaching Assistant Eilene Tomkins-Flanagan for the support!

## References

- 1.Gravelsins, L., Duncan, K., Einstein, G. (2021) Do Oral Contraceptives Affect Young Women's Memory? Dopamine-Dependent Working Memory Is Influenced by COMT Genotype, but Not Time of Pill Ingestion. PLOS ONE, vol. 16, no. 6. p. e0252807, <https://doi.org/10.1371/journal.pone.0252807>
- 2.Gravelsins, L., Duncan, K., Einstein, G. (2021) N-back data (Version #1) [Data set]. OSF.io <https://doi.org/10.17605/OSF.IO/35U6H>

# COGNITIVE CONTROL IN BILINGUALS AND MONOLINGUALS

TIANNA GABRIEL , SAFI KONE, DYA MAVARANI, JADA RICHARD

Carleton University, Department of Cognitive Science



## INTRODUCTION

Our research project replicates and extends the findings of Blumenfeld and Marian (2014).

- Purpose: provide a deeper understanding of bilingualism and related cognitive / language processes for practical implications and further research

Bilinguals engage in linguistic competition, which has been theorized to strengthen inhibitory control processes compared to monolinguals. In this study, two conflict types are examined:

- stimulus-stimulus conflict: when two features of a stimulus compete at the representational level, → measured through the Stroop task
- stimulus-response conflict: when a stimulus automatically triggers a motor response conflicting with a task's response rule. → measured through the Simon task

### Predictions / Hypothesis

- Bilinguals should show an advantage on the Stroop task through faster response times compared to monolinguals
- Language groups are hypothesized to affect Simon RTs and Stroop RTs
- Monolinguals should show no difference in performance across both task types
- The bilingual advantage should result from the cognitive mechanism embedded in the Stroop task ( stimulus-stimulus conflict), which they engage in more

## METHODS & MATERIALS

### Participants

- 21 university students, ages 18-22 ( $M$  age = 20)
- 61.9% female, 38.1% male
- Participants classified as: monolinguals (English), bilinguals (French, English) and multilingual (French, English and another given language).

### Materials

- Laptops used to record reaction times
- LexTALE-FR test
  - measured French proficiency
  - used to classify bilinguals vs monolinguals
- Stroop Task
  - Measured stimulus-stimulus inhibition
  - Participants identified the ink colour while ignoring the word
- Simon Task
  - measures stimulus-response inhibition
  - Participants responded to colour of the circle while ignoring it's location on the screen



## PROCEDURE

- Participants were recruited by email, social media or in person
- Participants completed:
  - 1. Demographic questionnaire
  - 2. LexTALE-FR language proficiency test
- Participants then completed two tasks:
  - Stroop Task
  - Simon Task
- Practice trials were provided before each task
- Reaction time and accuracy were recorded, but reaction time was what was used to test inhibitory control
- Total participation time: ~ 8-15 minutes



## RESULTS

Our results suggested no significant result between the bilinguals and monolinguals on either the Stroop ( $p = .469$ ) or Simon ( $p = .977$ ) tasks according to an independent samples t-test. Large standard deviation across scores suggests high variability across the scores. As seen in Figure 1, the scores of the Stroop Task were not very different, and the same can be said about the Simon Task in figure 2

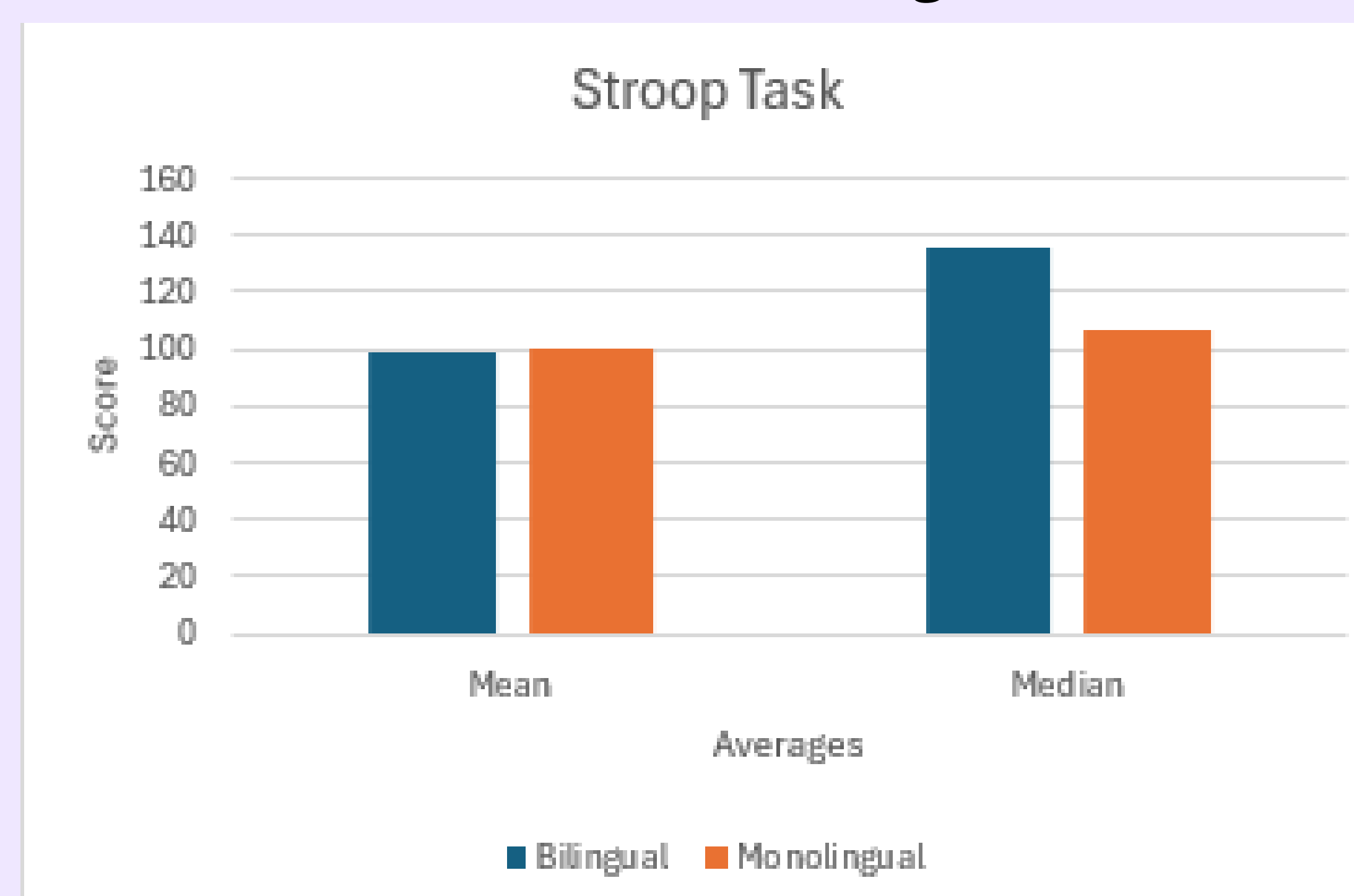
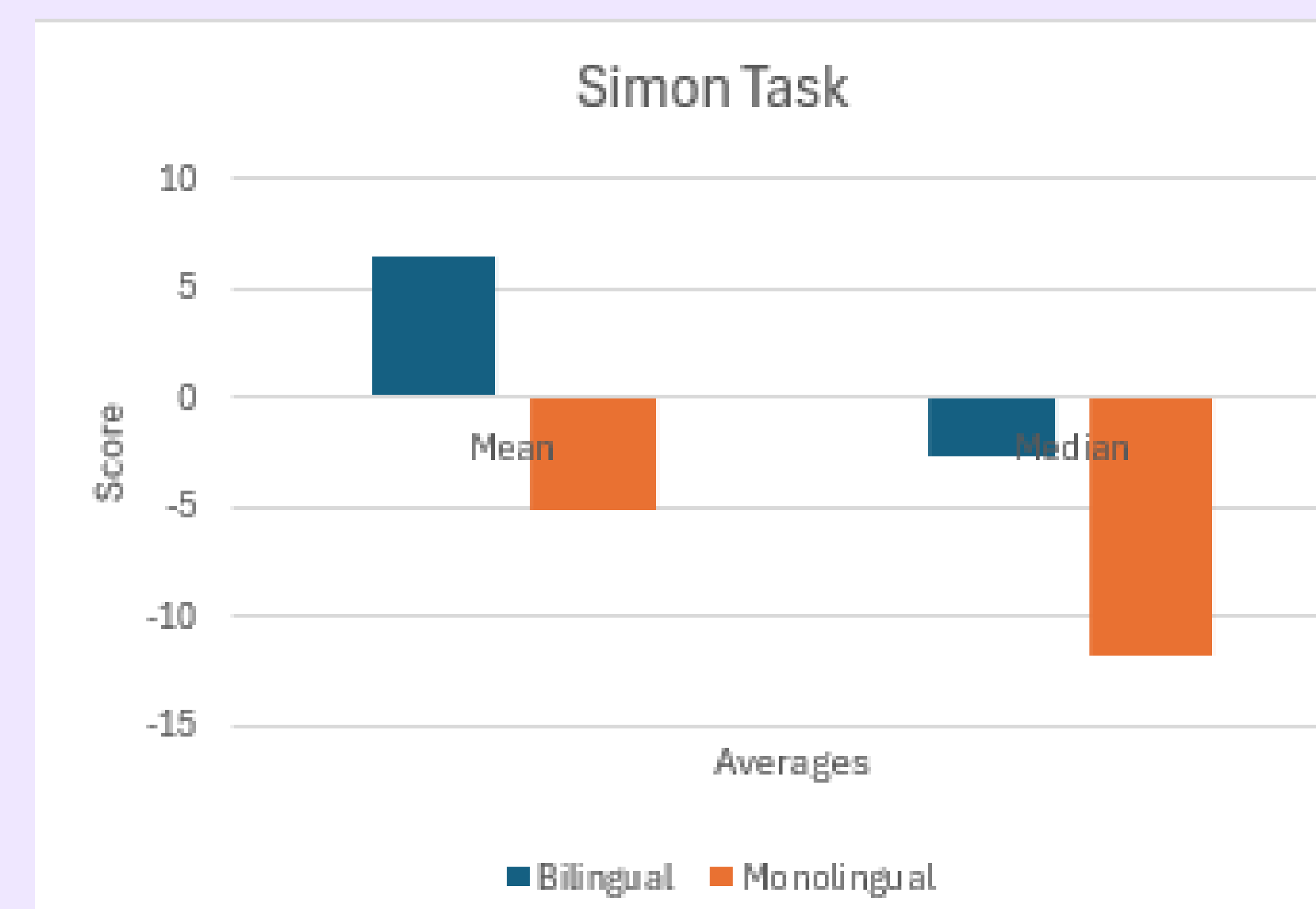


Figure 1: Means and Medians in the Stroop test

Figure 2: Means and Medians in the Simon Task



## IMPLICATIONS

These findings suggest that the bilingual cognitive advantage in inhibitory control may not be as consistent as proposed, challenging previous conclusions [1].

The high variability across scores suggests individual differences may play a stronger role in task performance than language group membership alone.

These results highlight the importance of precise participant classification, controlled testing environments, and sufficient trial counts for reliable and replicable findings in bilingual cognitive research.

## CONCLUSIONS

Bilingual cognitive advantage:

- Not a straightforward or universally replicable phenomenon. Further investigation is needed before firm conclusions can be drawn

Future research:

- Should operationalize bilingualism more precisely, use larger samples, standardize testing conditions, and explore variables such as language switching frequency and working memory capacity

Broader significance:

- Understanding when and how bilingualism strengthens cognitive control has implications for cognitive science, education, and our understanding of how language shapes the mind

## REFERENCES

- [1] Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus-Stimulus inhibition. *Bilingualism: Language and Cognition*, 17(3), 610-629. <https://doi.org/10.1017/S1366728913000564>

# Effect of Encoding Level and Test Type on False Memories

By: Daisy Hagens, Taylor Moon, and Olivia Schieber-Mule

## Introduction

### Deese-Roediger-McDermott (DRM) Paradigm:

Participants study a list of semantically similar words, used to test memory recall [1].

**Critical lure:** Semantically similar word not present in the DRM list [2].

**Research Question:** How does encoding level (deep vs. shallow) and test type (recognition vs. recall) impact memory accuracy and false memory rate?

**Hypothesis:** Higher false memory rate reported in deep-encoding, compared to shallow-encoding and higher false memory rate in recognition testing, compared to recall.

## Methodology

**Participants:** 32 adults aged 19-29 (M = 21.9, SD = 2.40), 20 females and 12 males, fluent English speakers.

### Materials:

Cold	Mountain
Hot, Snow, Warm, Winter, Ice, Wet, Frigid, Chilly, Heat, Weather, Freeze, Air, Shiver, Arctic, Frost	Hill, Valley, Climb, Summit, Top, Molehill, Peak, Plain, Glacier, Goat, Bike, Climber, Range, Steep, Ski

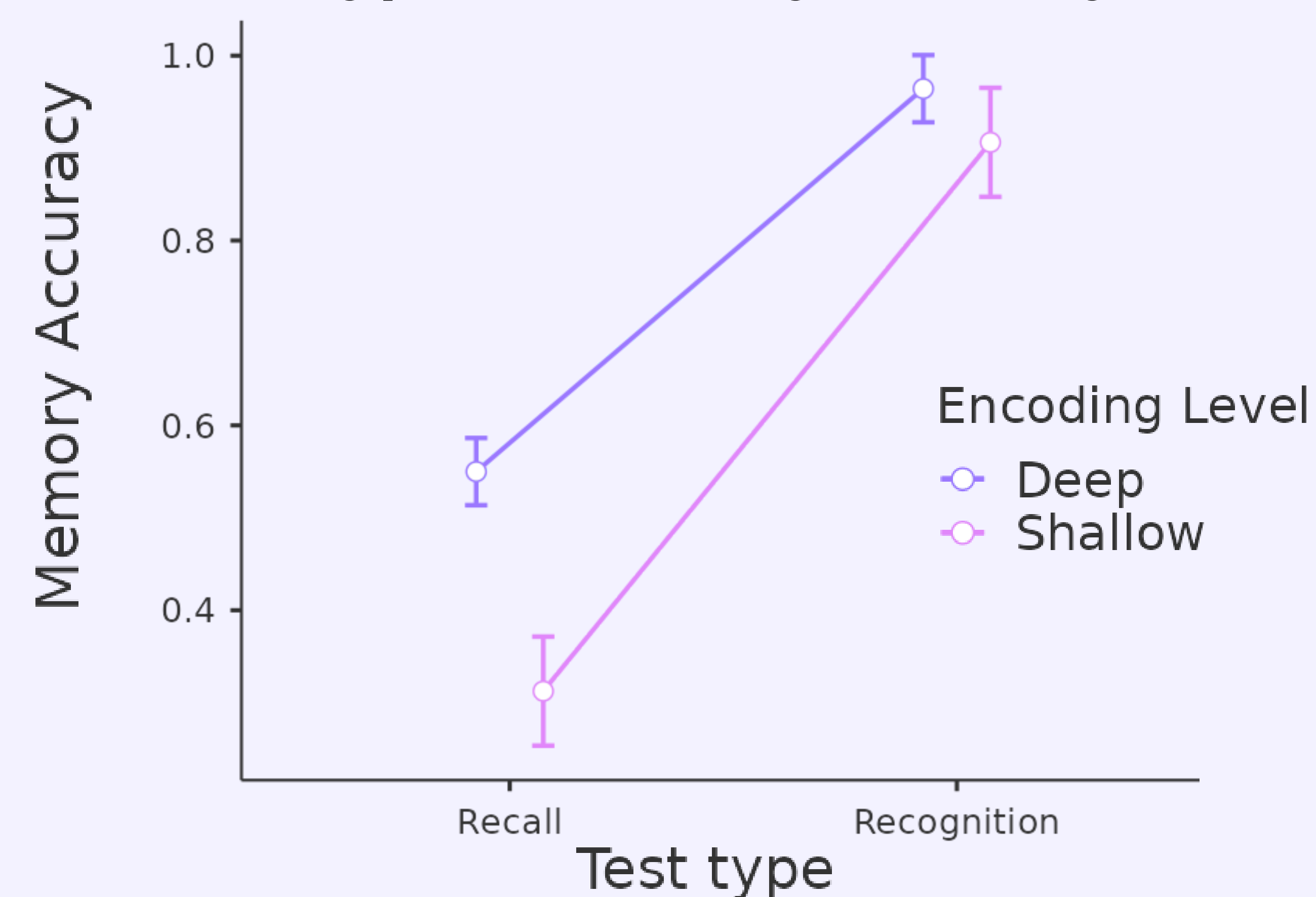
- Deep Processing: Daily Use 1 (not often) to 5 (often)
- Shallow Processing: Letter Count

### Procedure:

1. Study List (Shallow or Deep)
2. Recall or Recognition Task  
\*\*\* Distractor Task \*\*\*
3. Study Second List (Shallow or Deep)
4. Recall or Recognition Task

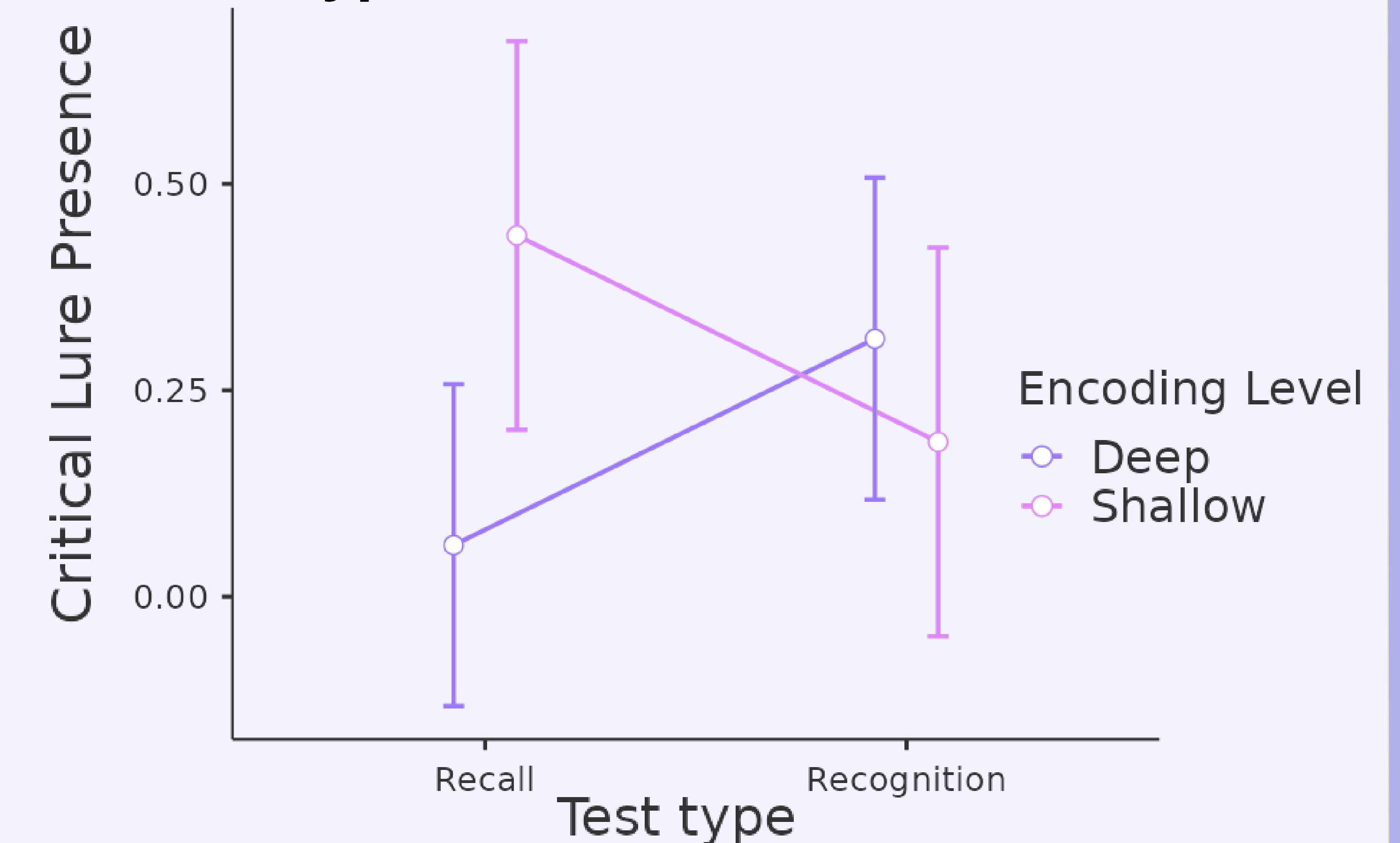
## Findings

### Interaction of Encoding Level and Test Type on Memory Accuracy



- Both Encoding Level and Test Type have significant main effects on memory accuracy.
- The Encoding Level × Test Type interaction on memory accuracy is significant.

### Interaction of Encoding Level and Test Type on Critical Lure Presence



- Neither Encoding Level nor Test Type had a significant main effect on critical lure presence.
- The Encoding Level × Test Type interaction on memory accuracy is significant.

## Discussion

These findings suggest that memory accuracy is higher on recognition tests compared to recall tests, and further increased during deep encoding. Test type and encoding level did not significantly affect critical lure presence, though the interaction meant that critical lure presence was highest on the recall-shallow condition and recognition-deep condition.

**Limitations:** Semantically similar words not included on the original list were also recalled, but we did not analyze other critical lures reported in the recall tests for consistency with the recognition test.

## References

- [1] Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17-22. <https://doi.org/10.1037/h0046671>
- [2] Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803-814. <https://doi.org/10.1037/0278-7393.21.4.803>



**Carleton University**

Department of Cognitive Science